



# Lipidomic and Genomic Investigation of Mahoney Lake, B.C.

## Citation

Bovee, Roderick. 2014. Lipidomic and Genomic Investigation of Mahoney Lake, B.C.. Doctoral dissertation, Harvard University.

## Permanent link

<http://nrs.harvard.edu/urn-3:HUL.InstRepos:11745724>

## Terms of Use

This article was downloaded from Harvard University's DASH repository, and is made available under the terms and conditions applicable to Other Posted Material, as set forth at <http://nrs.harvard.edu/urn-3:HUL.InstRepos:dash.current.terms-of-use#LAA>

## Share Your Story

The Harvard community has made this article openly available.  
Please share how this access benefits you. [Submit a story](#).

[Accessibility](#)

**Lipidomic and Genomic Investigation of Mahoney Lake, B.C.**

A dissertation presented  
by  
Roderick Bovee  
to  
The Department of Earth and Planetary Sciences

in partial fulfillment of the requirements  
for the degree of  
Doctor of Philosophy  
in the subject of  
Earth and Planetary Sciences

Harvard University  
Cambridge, Massachusetts

December, 2013

© 2013 – Roderick Bovee  
All rights reserved.

## **Lipidomic and Genomic Investigation of Mahoney Lake, B.C.**

### **Abstract**

Photic-zone euxinia (PZE) is associated with several times in Earth's history including Phanerozoic extinction events and long parts of the Proterozoic. One of the best modern analogues for extreme PZE is Mahoney Lake in British Columbia, Canada where a dense layer of purple sulfur bacteria separate the oxic mixolimnion from one of the most sulfidic monimolimnions in the world. These purple sulfur bacteria are known to produce the carotenoid okenone. Okenone's diagenetic product, okenane, has potential as a biomarker for photic-zone euxinia, so understanding its production and transport is important for interpreting the geologic record. In the following dissertation, I examine Mahoney Lake with a multi-proxy approach. I use lipid biomarkers to understand organic matter production burial in the lake and find strong evidence of lateral transport of organic matter from shoreline microbial mats to the lake-bottom sediments. I also find evidence of okenone production in these shoreline mats and a carotenoid previously unreported in the environment, Thiothece-484, associated with the okenone synthetic pathway. Finally, I develop a new bioinformatics method to examine high-throughput metagenomic data and use this method to start understanding how the metabolic and lipid synthetic pathways of microbial communities in the lake are associated with each other.

## TABLE OF CONTENTS

<b>Abstract</b>	iii
<b>Acknowledgments</b>	v
<b>Chapter 1 – Introduction</b>	1
<b>Chapter 2 – Fatty Acids</b>	14
<b>Chapter 3 – Pigments</b>	46
<b>Chapter 4 – Linking Biomarkers to Genomic Data</b>	86
<b>Appendix 1 – Data Presented in Chapters 1 &amp; 2</b>	132

## ACKNOWLEDGEMENTS

I have so much gratitude for my advisor, Ann Pearson, for all the knowledge she's given me and all the time she's spent on me. There have been many setbacks on the route to this dissertation, but Ann's optimism and drive helped push me through to the end. Thank you also to all the current and former members of my committee—Dave Johnston, Francis MacDonald, Colleen Hansel, Miaki Ishii, and Jerry Mitrovica—who have given so much of their time and support. I am particularly indebted to two of my former mentors, Garry Anderson (Saint Cloud State University) and Dallas Hinrichs (Mayo Clinic), whose guidance inspired me down this path. And without the administrative support of Sarah Colgan, Olga Kolas, Maryorie Grande, Chenoweth Moffatt, and Bridget Mastandrea, none of this would even be possible.

A huge thank you to Susan Carter, for sharing all of her laboratory expertise and for helping to fix all the small (and large!) things that inevitably go wrong in lab. Pearson lab members Felisa Wolfe-Simon, Wiebke Mohr, Tiantian Tang, and Sarah Sattin taught me so much about molecular biology in general and proteins specifically. Thank you also to post-docs Lindsay Hays, Stephanie Kusch, Evelyn Zeiler, and Alex Bradley and undergrads Clarmyra Hayes, Yelun Qin, Leah Tsao, and Lyle Nelson for everything they've taught me from culturing to chemistry.

The graduate students in the EPS department are terrific and the wonderful departmental community here is something I'm really grateful for. In particular, thanks to Rita Parai, my field-trip co-leader, and all the other “true” geologists in the department for helping me not stray too

far from my roots. So many thanks to my fellow Peason Lab graduate students, Meytal Higgins, Hilary Close, Yundan Pi, Sarah Hurley, and Elise Wilkes, and our honorary lab member Wil Leavitt, for all of their camaraderie and support.

Thanks to my roommates at 14A Pitman—Cassie, Kristin, and Tor—and my roommates at 58 Ibbetson—Hannah, Jamie, Taylor, and Marena—for their friendships and for introducing me to so many things broader than science.

I really could have done none of this without the support of my family. I am so grateful for everything that my sister Regan, my niece Sophia, my brother-in-law Nick, my grandfather Walt, my grandmother Joan, my mother Teresa and my father John have given me. And last, but certainly not least, so many thanks to Clarissa: your support and patience over the last two years have been such a gift.

## CHAPTER 1

### Photic-Zone Euxinia in the Geologic Record

#### *Photic Zone Euxinia*

Atmospheric and marine oxygen levels were lower than the present day through most of the Precambrian (Canfield, 2005; Holland, 2006; Hayes and Waldbauer, 2006) and marine oxygen levels have dropped occasionally during the Phanerozoic (Jenkyns, 1980; Isozaki, 1997).

Tracking both long-term secular changes and short term variability in marine oxygen levels requires understanding anoxia and the records it leaves behind. In modern marine environments, there are two principal styles of persistent ( $> 1$  year) anoxia: oxygen minimum zones (OMZs) where anoxia is driven by oxygen depletion through remineralization of high biological productivity and restricted basins where lack of circulation creates a lack of oxygen replenishment. It is not clear if past episodes of marine anoxia more resembled the productivity-driven style (Pedersen and Calvert, 1990) or the circulation-driven styles (Degens and Stoffers, 1976; Hay, 2008) or some combination thereof (Schlanger and Jenkyns, 1976). One particularly important factor in interpreting paleoanoxia is vertical extent; have euxinic marine waters expanded into the photic zone, so-called photic zone euxinia (PZE).

The best current proxies for PZE come from two groups of sulfide-oxidizing photoautotrophs: the Green Sulfur Bacteria (GSB) and the Purple Sulfur Bacteria (PSB). GSB produce the diagnostic carotenoids isorenieratene or chlorobactene and bacteriochlorophylls *c*, *d*, and *e* while PSB produce okenone and bacteriochlorophyll *a* (Blankenship et al., 1995). Marine populations



of GSB have only been reported from modern restricted basins and not OMZs. The classic locale for modern GSB is the Black Sea. Here, GSB-derived pigments are present in the chemocline of the Black Sea (Repeta et al., 1989) and in the basinal sediments (Repeta, 1993; Sinninghe Damsté et al., 1993). Consistent with its source from photoautotrophs, isorenieratene has not been found in deeper, subphotic anoxic basins such as Cariaco (e.g. Wakeham et al., 2012).

Purple Sulfur Bacteria, on the other hand, are generally undetectable in both restricted basins and OMZs, but common in some meromictic lakes. Hundred to thousand year records of okenone have been reported from the sediments of several such lakes (Guilizzoni et al., 1986; Züllig 1986; Overmann et al., 1993; Itoh et al., 2003; Rogozin et al., 2011) and some isolated fjords (Smittenberg et al., 2004). Okenone, chlorobactene, and isorenieratene all have saturated diagenetic derivatives that can be geologically preserved: okenane, chlorobactane, and isorenieratane respectively (Summons and Powell, 1986; Schaeffer et al., 1997).

Modern OMZs still contain measurable nitrate and sometimes oxygen, so sulfate reduction is not energetically favorable and sulfide levels are below detection. Low sulfide prevents the growth of large populations of PSB and GSB (although there may be “cryptic” sulfide fluxes; Canfield et al., 2010). Instead, cyanobacteria like *Prochlorococcus* are known to grow at the top of the OMZ (~100 m depth) in both the Eastern Tropical North Pacific (ETNP) and Arabian Sea (Goericke et al., 2000; Ma et al., 2009). Given the light and nutrient gradients, it is likely that anoxygenic photosynthesis would be viable if sulfide were present. Consistent with this hypothesis, transcripts of sulfur-metabolizing genes presumed to be from purple and green sulfur bacteria

have been reported from the Eastern Tropical South Pacific (ETSP) OMZ (Stewart et al., 2012) although no reports of okenone, isorenieratene, or chlorobactene have been reported from any OMZ.

#### *Time Periods with PZE*

Much as with the distributions of their precursors in the modern ocean, isorenieratane is much more common than okenane in the Phanerozoic record. The first use of isorenieratane as a geologic proxy for PZE was on rocks derived from the Devonian epeiric seas of North America (Summons and Powell, 1986; Hartgers et al., 1993; Behrens et al., 1998; Brown and Kenig, 2004). Later work expanded observations to Mesozoic Ocean Anoxic Events (OAEs) where isorenieratene and/or other biomarkers for PZE were found in rocks at or temporally near the Jurassic OAEs (Van Kaam-Peters and Sinninghe Damsté, 1997; Kenig et al., 2004; Schwark and Frimmel 2004; Pancost et al., 2004), OAE1b (Heimhofer et al., 2008; Tzortzaki et al., 2013), OAE 2 (Sinninghe Damsté and Köster, 1998; Kuypers et al., 2002; 2004; Simons et al., 2003; Kolonic et al., 2005; van Bentum et al., 2009; Sepúlveda et al., 2009) and OAE3 (Wagner et al., 2004). All these Mesozoic observations were made in the Atlantic or closely related marginal basins; similar geographic restriction is present in Cenozoic PZE observations. During this era, isorenieratene has been reported in the Mediterranean from Messinian marls (Kohnen et al., 1992; Keely et al., 1995; Kenig et al., 1995; Putschew et al., 1998) and Pliocene sapropels (Passier et al., 1999; Menzel et al., 2002) and in the Arctic ocean during the Paleocene/Eocene Thermal Maximum (Sluijs et al., 2006).

The only time during the Phanerozoic when ocean-wide photic-zone euxinia may have occurred is at the Permo-Triassic transition. Here, isorenieratane and aryl isoprenoids have been reported from the two major oceans: the Paleotethys Ocean (Grice et al., 2005a; 2005b; Cao et al., 2009) and the Panthalassic Ocean (Hays et al., 2007; Hays et al., 2012), in addition to inland seas (Grice et al., 1996a). Malemides that are believed to be derived from GSB bacteriochlorophylls have also been reported from some of these sections (Grice et al., 1996b; Pancost et al., 2002). Despite the prevalence of anoxia during this time period, it is not clear that anoxia is necessarily synchronous with the extinction (Nielsen et al., 2010).

#### *Issues with Isorenieratane as a Marker for PZE*

There may be several issues with using isorenieratane as a biomarker for photic zone euxinia. Isorenieratane may not be a diagnostic biomarker for just Chlorobi; it has long been recognized that Actinomycetes produce isorenieratane and may also be a source of it in aerobic environments (Keely and Maxwell, 1993). Additionally, isorenieratane can be formed during in-sediment cyclization of acyclic carotenoids like  $\beta$ -carotene (Koopmans et al., 1996) which are produced by a wide taxonomic variety of plants and microorganisms. Both of these shortcomings can be ameliorated by isotopic measurements of isorenieratane or its derivatives because GSB use the reverse tricarboxylic acid (rTCA) carbon fixation pathway which fractionates carbon less than the normal Rubisco pathway resulting in isotopically heavier isorenieratane (Sirevåg et al., 1977).

Although they are obligate anaerobes, GSB do not require the presence of sulfide: populations

have also been reported from the  $\text{Fe}^{2+}$  rich and sulfide poor chemocline of Lake Matano (Crowe et al., 2008). Under ferruginous conditions, GSB-derived products would still be diagnostic for photic zone anoxia, but not PZE. Lastly, having a good understanding of provenance to rule out lateral inputs is important for all biomarkers. Isorenieratane has been found in some well-oxygenated, open-ocean sediments as a detrital input (Rosell-Melé, et al., 1997) so as with all biomarker measurements, geologic context is important in their interpretation.

#### *Okenane as a Alternate PZE Biomarker*

In contrast to the rich biomarker record of isorenieratane and other derivatives of GSB, there are only two observations of okenane before the Quaternary. The first is from 25-65 Ma saline lacustrine deposits from China (Zhang et al., 2011). These deposits are likely similar to the present-day lakes where okenone has been reported. The second is from the marine Barney Creek 1.6 Ga Mesoproterozoic deposits in the McArthur basin in Australia (Brocks et al., 2005; 2008). These rocks were initially interpreted as shallow saline lake deposits, but are now believed to have formed in an open marine setting (Rawlings, 1999). If this newer interpretations hold, this is a radically different environment from where PSB are found at present.

As PSB live much shallower in the water column, the presence of okenane implies much shallower sulfidic conditions than the deeper conditions isorenieratane would be found in (Montesinos et al., 1983). If the chemocline is shallow enough for purple sulfur bacteria to grow, it should lie well within the mixed layer and large amounts of sulfide should escape to the atmosphere. Additionally, deep waters would presumably be very euxinic in such a scenario.

This is consistent with initial models where sulfur isotopes pointed towards the 1.8-1.0 Ga oceans as being superficially oxic with euxinia at depth (Canfield, 1998). However more recent iron speciation and other data points towards zones of ferruginous deep water and sulfidic OMZs during the Late Paleoproterozoic (Poulton et al., 2010) and Mesoproterozoic (Planavsky et al., 2011). Further confirming the unusualness of open-ocean Mesoproterozoic okenone, rocks from a 1.1 Ga Mesoproterozoic epeiric basin contained aryl isoprenoids derived from isorenieratene, but not okenane (Blumenberg et al., 2012).

Similar conditions where sulfide fluxes were high enough for atmospheric escape have been posited for the Permo-Triassic (Kump et al., 2005; Meyer et al., 2008) although okenane has not been reported from anywhere during this time period. Transgressive rocks from the Permo-Triassic similar to those from the Barney Creek do not contain okenone or its derivatives, only aryl isoprenoids derived from isorenieratene and/or chlorobactene (Nabbefeld et al., 2010). At least two alternate explanations for the presence of okenane in the Mesoproterozoic are possible: either the depositional environment of the Barney Creek Formation more closely resembled modern salinity stratified systems (Lyons et al., 2009) or okenone is not a reliable marker for extremely shallow PZE.

Resolving this requires a better understanding of the conditions under which PSB live, produce okenone, and are buried. One site with an exceptionally dense layer of PSB is the chemocline of meromictic Mahoney Lake in British Columbia, Canada (Northcote and Hall, 1969; Overmann et al., 1991). Here, I investigate the biological and sedimentary processes in Mahoney Lake using

complementary low molecular weight lipid (Chapter 2), pigment (Chapter 3) and bioinformatic (Chapter 4) approaches.

## References

Behrens A, Wilkes H, Schaeffer P, Clegg H, Albrecht P (1998) Molecular characterization of organic matter in sediments from the Keg River formation (Elk Point group), western Canada sedimentary basin. *Organic Geochemistry*, **29**(8), 1905–1920.

Blankenship RE, Madigan MT, Bauer CCE (1995) *Anoxygenic photosynthetic bacteria*. Springer, Vol. 2.

Blumenberg M, Thiel V, Riegel W, Kah LC, Reitner J (2012) Biomarkers of black shales formed by microbial mats, Late Mesoproterozoic (1.1 Ga) Taoudeni Basin, Mauritania. *Precambrian Research*, **196–197**, 113–127.

Brocks JJ, Love GD, Summons RE, Knoll AH, Logan GA, Bowden SA (2005) Biomarker evidence for green and purple sulphur bacteria in a stratified Palaeoproterozoic sea. *Nature*, **437**(7060), 866–870.

Brocks JJ, Schaeffer P (2008) Okenane, a biomarker for purple sulfur bacteria (Chromatiaceae), and other new carotenoid derivatives from the 1640 Ma Barney Creek Formation. *Geochimica et Cosmochimica Acta*, **72**(5), 1396–1414.

Brown TC, Kenig F (2004) Water column structure during deposition of Middle Devonian–Lower Mississippian black and green/gray shales of the Illinois and Michigan Basins: a biomarker approach. *Palaeogeography, Palaeoclimatology, Palaeoecology*, **215**(1–2), 59–85.

Canfield D (1998) A new model for Proterozoic ocean chemistry. *Nature*, **396**(6710), 450.

Canfield DE (2005) The early history of atmospheric oxygen: homage to Robert M. Garrels. *Annu. Rev. Earth Planet. Sci.*, **33**, 1–36.

Canfield DE, Stewart FJ, Thamdrup B, De Brabandere L, Dalsgaard T, Delong EF, Revsbech NP, Ulloa O (2010) A Cryptic Sulfur Cycle in Oxygen-Minimum-Zone Waters off the Chilean Coast. *Science*, **330**(6009), 1375.

Cao C, Love GD, Hays LE, Wang W, Shen S, Summons RE (2009) Biogeochemical evidence for euxinic oceans and ecological disturbance presaging the end-Permian mass extinction event. *Earth and Planetary Science Letters*, **281**(3–4), 188–201.

Crowe SA, Jones C, Katsev S, Magen C, O'Neill AH, Sturm A, Canfield DE, Haffner GD, Mucci A, Sundby B, Fowle DA (2008) Photoferrotrophs thrive in an Archean Ocean analogue. *Proceedings of the National Academy of Sciences*, **105**(41), 15938–15943.

Degens ET, Stoffers P (1976) Stratified waters as a key to the past. *Nature*, **263**(5572), 22–27.

Goericke R, Olson R., Shalapyonok A (2000) A novel niche for *Prochlorococcus* sp. in low-light suboxic environments in the Arabian Sea and the Eastern Tropical North Pacific. *Deep Sea Research Part I: Oceanographic Research Papers*, **47**(7), 1183–1205.

Grice K, Schaeffer P, Schwark L, Maxwell JR (1996a) Molecular indicators of palaeoenvironmental conditions in an immature Permian shale (Kupferschiefer, Lower Rhine Basin, north-west Germany) from free and S-bound lipids. *Organic Geochemistry*, **25**(3–4), 131–147.

Grice K, Gibbison R, Atkinson JE, Schwark L, Eckardt CB, Maxwell JR (1996b) Maleimides (1H-pyrrole-2,5-diones) as molecular indicators of anoxygenic photosynthesis in ancient water columns. *Geochimica et Cosmochimica Acta*, **60**(20), 3913–3924.

Grice K, Twitchett RJ, Alexander R, Foster CB, Looy C (2005a) A potential biomarker for the Permian–Triassic ecological crisis. *Earth and Planetary Science Letters*, **236**(1–2), 315–321.

Grice K, Cao C, Love GD, Twitchett RJ, Bottcher M, Grosjean E, Summons RE, Turgeon SC, Dunning W, Jin Y (2005b) Photic Zone Euxinia During the Permian-Triassic Superanoxic Event. *Science*, **307**(5710), 706–709.

Guilizzoni P, Lami A, Ruggiu D, Bonomi G (1986) Stratigraphy of specific algal and bacterial carotenoids in the sediments of Lake Varese (N. Italy). *Hydrobiologia*, **143**(1), 321–325.

Hartgers WA, Damsté JSS, Koopmans MP, de Leeuw JW (1993) Sedimentary evidence for a diaromatic carotenoid with an unprecedented aromatic substitution pattern. *Journal of the Chemical Society, Chemical Communications*, (23), 1715–1716.

Hayes JM, Waldbauer JR (2006) The carbon cycle and associated redox processes through time. *Philosophical Transactions of the Royal Society B: Biological Sciences*, **361**(1470), 931–950.

Hays LE, Beatty T, Henderson CM, Love GD, Summons RE (2007) Evidence for photic zone euxinia through the end-Permian mass extinction in the Panthalassic Ocean (Peace River Basin, Western Canada). *Palaeoworld*, **16**(1), 39–50.

Hays LE, Grice K, Foster CB, Summons RE (2012) Biomarker and isotopic trends in a Permian–Triassic sedimentary section at Kap Stosch, Greenland. *Organic Geochemistry*, **43**, 67–82.

Hay WW (2008) Evolving ideas about the Cretaceous climate and ocean circulation. *Cretaceous Research*, **29**(5–6), 725–753.

Heimhofer U, Hesselbo SP, Pancost RD, Martill DM, Hochuli PA, Guzzo JVP (2008) Evidence for photic-zone euxinia in the Early Albian Santana Formation (Araripe Basin, NE Brazil). *Terra Nova*, **20**(5), 347–354.

Holland HD (2006) The oxygenation of the atmosphere and oceans. *Philosophical Transactions of the Royal Society B: Biological Sciences*, **361**(1470), 903–915.

Isozaki Y (1997) Permo-Triassic Boundary Superanoxia and Stratified Superocean: Records from Lost Deep Sea. *Science*, **276**(5310), 235–8.

Itoh N, Tani Y, Nagatani T, Soma M (2003) Phototrophic activity and redox condition in Lake Hamana, Japan, indicated by sedimentary photosynthetic pigments and molybdenum over the last ~250 years. *Journal of Paleolimnology*, **29**(4), 403–422.

Jenkyns HC (1980) Cretaceous anoxic events: from continents to oceans. *Journal of the Geological Society*, **137**(2), 171–188.

Keely BJ, Blake SR, Schaeffer P, Maxwell JR (1995) Distributions of pigments in the organic matter of marls from the Vena del Gesso evaporitic sequence. *Organic Geochemistry*, **23**(6), 527–539.

Keely BJ, Maxwell JR (1993) The Mulhouse basin: evidence from porphyrin distributions for water column anoxia during deposition of marls. *Organic Geochemistry*, **20**(8), 1217–1225.

Kenig F, Damsté JS, Frewin NL, Hayes JM, De Leeuw JW (1995) Molecular indicators for palaeoenvironmental change in a Messinian evaporitic sequence (Vena del Gesso, Italy). II: High-resolution variations in abundances and <sup>13</sup>C contents of free and sulphur-bound carbon skeletons in a single marl bed. *Organic Geochemistry*, **23**(6), 485–526.

Kenig F, Hudson JD, Damsté JSS, Popp BN (2004) Intermittent euxinia: Reconciliation of a Jurassic black shale with its biofacies. *Geology*, **32**(5), 421–424.

Kohnen ME, Schouten S, Damsté JSS, de Leeuw JW, Merritt DA, Hayes JM (1992) Recognition of paleobiochemicals by a combined molecular sulfur and isotope geochemical approach. *Science*, **256**(5055), 358–362.

Kolonis S, Wagner T, Forster A, Damsté JS., Walsworth-Bell B, Erba E, Turgeon S, Brumsack HJ, Chellai EI, Tsikos H, others (2005) Black shale deposition on the northwest African Shelf during the Cenomanian/Turonian oceanic anoxic event: Climate coupling and global organic carbon burial. *Paleoceanography*, **20**(1).

Koopmans MP, Köster J, Van Kaam-Peters HME, Kenig F, Schouten S, Hartgers WA, de Leeuw JW, Sinninghe Damsté JS (1996) Diagenetic and catagenetic products of isorenieratene:



Molecular indicators for photic zone anoxia. *Geochimica et Cosmochimica Acta*, **60**(22), 4467–4496.

Kump LR, Pavlov A, Arthur MA (2005) Massive release of hydrogen sulfide to the surface ocean and atmosphere during intervals of oceanic anoxia. *Geology*,

Kuypers MMM, Blokker P, Hopmans EC, Kinkel H, Pancost RD, Schouten S, Sinninghe Damsté JS (2002) Archaeal remains dominate marine organic matter from the early Albian oceanic anoxic event 1b. *Palaeogeography, Palaeoclimatology, Palaeoecology*, **185**(1-2), 211–234.

Kuypers MMM, Lourens LJ, Rijpstra WIC, Pancost RD, Nijenhuis IA, Sinninghe Damsté JS (2004) Orbital forcing of organic carbon burial in the proto-North Atlantic during oceanic anoxic event 2. *Earth and Planetary Science Letters*, **228**(3–4), 465–482.

Lyons TW, Reinhard CT, Scott C (2009) Redox Redux. *Geobiology*, **7**(5), 489–494.

Ma Y, Zeng Y, Jiao N, Shi Y, Hong N (2009) Vertical distribution and phylogenetic composition of bacteria in the Eastern Tropical North Pacific Ocean. *Microbiological Research*, **164**(6), 624–633.

Menzel D, Hopmans EC, van Bergen PF, de Leeuw JW, Sinninghe Damsté JS (2002) Development of photic zone euxinia in the eastern Mediterranean Basin during deposition of Pliocene sapropels. *Marine Geology*, **189**(3–4), 215–226.

Meyer KM, Kump LR, Ridgwell A (2008) Biogeochemical controls on photic-zone euxinia during the end-Permian mass extinction. *Geology*, **36**(9), 747–750.

Montesinos E, Guerrero R, Abella C, Esteve I (1983) Ecology and Physiology of the Competition for Light Between *Chlorobium limicola* and *Chlorobium phaeobacteroides* in Natural Habitats. *Applied and Environmental Microbiology*, **46**(5), 1007–1016.

Nabbefeld B, Grice K, Twitchett RJ, Summons RE, Hays L, Böttcher ME, Asif M (2010) An integrated biomarker, isotopic and palaeoenvironmental study through the Late Permian event at Lusitaniadalen, Spitsbergen. *Earth and Planetary Science Letters*, **291**(1–4), 84–96.

Nielsen JK, Shen Y, Piasecki S, Stemmerik L (2010) No abrupt change in redox condition caused the end-Permian marine ecosystem collapse in the East Greenland Basin. *Earth and Planetary Science Letters*, **291**(1–4), 32–38.

Northcote TG, Halsey TG (1969) Seasonal Changes in the Limnology of Some Meromictic Lakes in Southern British Columbia. *Journal of the Fisheries Research Board of Canada*, **26**(7), 1763–1787.

Overmann J, Beatty JT, Hall KJ, Pfennig N, Northcote TG (1991) Characterization of a dense, purple sulfur bacterial layer in a meromictic salt lake. *Limnology and Oceanography*, **36**(3), 846–859.

Overmann J, Sandmann G, Hall KJ, Northcote TG (1993) Fossil carotenoids and paleolimnology of meromictic Mahoney Lake, British Columbia, Canada. *Aquatic Sciences*, **55**(1), 31–39.

Pancost RD, Crawford N, Magness S, Turner A, Jenkyns HC, Maxwell JR (2004) Further evidence for the development of photic-zone euxinic conditions during Mesozoic oceanic anoxic events. *Journal of the Geological Society*, **161**(3), 353–364.

Pancost RD, Crawford N, Maxwell JR (2002) Molecular evidence for basin-scale photic zone euxinia in the Permian Zechstein Sea. *Chemical Geology*, **188**(3–4), 217–227.

Passier HF, Bosch H-J, Nijenhuis IA, Lourens LJ, Böttcher ME, Leenders A, Damsté JSS, de Lange GJ, Leeuw JW (1999) Sulphidic Mediterranean surface waters during Pliocene sapropel formation. *Nature*, **397**(6715), 146–149.

Pedersen TF, Calvert SE (1990) Anoxia vs productivity; what controls the formation of organic-carbon-rich sediments and sedimentary rocks? *AAPG Bulletin*, **74**(4), 454–466.

Planavsky NJ, McGoldrick P, Scott CT, Li C, Reinhard CT, Kelly AE, Chu X, Bekker A, Love GD, Lyons TW (2011) Widespread iron-rich conditions in the mid-Proterozoic ocean. *Nature*, **477**(7365), 448–451.

Poulton SW, Fralick PW, Canfield DE (2010) Spatial variability in oceanic redox structure 1.8 billion years ago. *Nature Geoscience*, **3**(7), 486–490.

Putschew A, Schaeffer P, Schaeffer-Reiss C, Maxwell JR (1998) Carbon isotope characteristics of the diaromatic carotenoid, isorenieratene (intact and sulfide-bound) and a novel isomer in sediments. *Organic Geochemistry*, **29**(8), 1849–1856.

Rawlings DJ (1999) Stratigraphic resolution of a multiphase intracratonic basin system: the McArthur Basin, northern Australia. *Australian Journal of Earth Sciences*, **46**(5), 703–723.

Repeta DJ (1993) A high resolution historical record of Holocene anoxygenic primary production in the Black Sea. *Geochimica et Cosmochimica Acta*, **57**(17), 4337–4342.

Repeta DJ, Simpson DJ, Jorgensen BB, Jannasch HW (1989) Evidence for anoxygenic photosynthesis from the distribution of bacterio-chlorophylls in the Black Sea. *Nature*, **342**(6245), 69–72.

Rogozin DY, Zykov VV, Kalugin IA, Daryin AV, Degermendzhy AG (2011) Carotenoids of phototrophic organisms in bottom sediments of meromictic Lake Shira (Siberia, Russia) as an indicator of past stratification. *Doklady Biological Sciences*, **439**(1), 228–231.

Rosell-Melé A, Maslin MA, Maxwell JR, Schaeffer P (1997) Biomarker evidence for “Heinrich” events. *Geochimica et Cosmochimica Acta*, **61**(8), 1671–1678.

Schaeffer P, Adam P, Wehrung P, Albrecht P (1997) Novel aromatic carotenoid

derivatives from sulfur photosynthetic bacteria in sediments. *Tetrahedron Letters*, **38**(48), 8413–8416.

Schlanger SO, Jenkyns HC (1976) Cretaceous oceanic anoxic events: causes and consequences. *Geologie en mijnbouw*, **55**(3-4), 179–184.

Schwark L, Frimmel A (2004) Chemostratigraphy of the Posidonia Black Shale, SW-Germany: II. Assessment of extent and persistence of photic-zone anoxia using aryl isoprenoid distributions. *Chemical Geology*, **206**(3–4), 231–248.

Sepúlveda J, Wendler J, Leider A, Kuss H-J, Summons RE, Hinrichs K-U (2009) Molecular isotopic evidence of environmental and ecological changes across the Cenomanian–Turonian boundary in the Levant Platform of central Jordan. *Organic Geochemistry*, **40**(5), 553–568.

Simons D-JH, Kenig F, Schröder-Adams CJ (2003) An organic geochemical study of Cenomanian-Turonian sediments from the Western Interior Seaway, Canada. *Organic Geochemistry*, **34**(8), 1177–1198.

Sinninghe Damsté JS, Köster J (1998) A euxinic southern North Atlantic Ocean during the Cenomanian/Turonian oceanic anoxic event. *Earth and Planetary Science Letters*, **158**(3–4), 165–173.

Sinninghe Damsté JS, Wakeham SG, Kohnen MEL, Hayes JM, de Leeuw JW (1993) A 6,000-year sedimentary molecular record of chemocline excursions in the Black Sea. *Nature*, **362**(6423), 827–829.

Sirevåg R, Buchanan BB, Berry JA, Troughton JH (1977) Mechanisms of CO<sub>2</sub> fixation in bacterial photosynthesis studied by the carbon isotope fractionation technique. *Archives of Microbiology*, **112**(1), 35–38.

Sluijs A, Schouten S, Pagani M, Woltering M, Brinkhuis H, Sinninghe Damsté JS, Dickens GR, Huber M, Reichert G-J, Stein R, Matthiessen J, Lourens LJ, Pedentchouk N, Backman J, Moran K (2006) Subtropical Arctic Ocean temperatures during the Palaeocene/Eocene thermal maximum. *Nature*, **441**(7093), 610–613.

Smittenberg R., Pancost R., Hopmans E., Paetzel M, Sinninghe Damsté J. (2004) A 400-year record of environmental change in an euxinic fjord as revealed by the sedimentary biomarker record. *Palaeogeography, Palaeoclimatology, Palaeoecology*, **202**(3–4), 331–351.

Stewart FJ, Ulloa O, DeLong EF (2012) Microbial metatranscriptomics in a permanent marine oxygen minimum zone. *Environmental Microbiology*, **14**(1), 23–40.

Summons RE, Powell TG (1986) Chlorobiaceae in Palaeozoic seas revealed by biological markers, isotopes and geology. *Nature*, **319**(6056), 763–765.

Tzortzaki E, Karakitsios V, Tsikos H (2013) Biomarker evidence for intermittent photic

zone euxinia in the Aptian–Albian organic sedimentary record from the Ionian Zone (Epirus, Greece). *Organic Geochemistry*, **64**, 84–93.

Van Bentum EC, Hetzel A, Brumsack H-J, Forster A, Reichart G-J, Sinninghe Damsté JS (2009) Reconstruction of water column anoxia in the equatorial Atlantic during the Cenomanian–Turonian oceanic anoxic event using biomarker and trace metal proxies. *Palaeogeography, Palaeoclimatology, Palaeoecology*, **280**(3–4), 489–498.

Van Kaam-Peters HM., Sinninghe Damsté JS (1997) Characterisation of an extremely organic sulphur-rich, 150 Ma old carbonaceous rock: palaeoenvironmental implications. *Organic Geochemistry*, **27**(7–8), 371–397.

Wagner T, Sinninghe Damsté JS, Hofmann P, Beckmann B (2004) Euxinia and primary production in Late Cretaceous eastern equatorial Atlantic surface waters fostered orbitally driven formation of marine black shales. *Paleoceanography*, **19**(3).

Wakeham SG, Turich C, Schubotz F, Podlaska A, Li XN, Varela R, Astor Y, Sáenz JP, Rush D, Sinninghe Damsté JS, Summons RE, Scranton MI, Taylor GT, Hinrichs K-U (2012) Biomarkers, chemistry and microbiology show chemoautotrophy in a multilayer chemocline in the Cariaco Basin. *Deep Sea Research Part I: Oceanographic Research Papers*, **63**, 133–156.

Zhang C, Zhang Y, Cai C (2011) Aromatic isoprenoids from the 25–65 Ma saline lacustrine formations in the western Qaidam Basin, NW China. *Organic Geochemistry*, **42**(7), 851–855.

Züllig H (1986) Carotenoids from plankton and photosynthetic bacteria in sediments as indicators of trophic changes in Lake Lobsigen during the last 14000 years. *Hydrobiologia*, **143**(1), 315–319.

## **CHAPTER 2**

### **Strong influence of the littoral zone on basinal sediments of a meromictic lake**

Photic zone euxinia (PZE) is an important source of primary production in stratified, sulfur-rich environments. The potential for export and burial of this material in sediments remains relatively understudied, despite being of fundamental importance to interpreting the geologic record of bulk total organic carbon (TOC) and individual lipid biomarkers. Here we report the relative concentrations and carbon isotope contents of lipid biomarkers from meromictic Mahoney Lake. The data indicate that all of the organic burial in the central basin sediments may derive from shoreline organic matter, and that the source of this material is from the littoral microbial community and/or detrital remains of terrigenous higher plants. Our results confirm that due to the strong density stratification and the intensive carbon and sulfur recycling pathways in the water column, there is minimal direct export of the sulfur-oxidizing planktonic community to depth. In contrast, material from the lake margins appears to traverse this same density gradient, travel down-slope, and become deposited in the deep basin. This suggests an important role for clastic and authigenic minerals in aiding the sedimentation and burial of terrigenous and mat-derived organic matter in euxinic systems.

#### **Introduction**

To examine processes controlling organic matter delivery to sediments in euxinic systems, we extracted and analyzed lipid biomarkers from Mahoney Lake, British Columbia, Canada. Mahoney Lake has an oligotrophic mixolimnion overlying a permanently-stratified hypolimnion that contains extremely high concentrations of sulfate and sulfide (Northcote and Hall, 1983).

The chemocline of the lake is stable at 7 m depth, where it supports the densest population of purple sulfur bacteria (PSB) (*Chromatiaceae*) ever measured (bacteriochlorophyll *a* > 20 mg l<sup>-1</sup>; Overmann et al., 1991). The dominant species in this layer had been believed to be the photoautotroph *Lamprocystis purpurea* (formerly *Amoebobacter purpureus*; Overmann et al., 1991), but companion analyses to this work now show that the major organism is a *Chromatiaceae* sp. closely related at the whole-genome level to *Thiorhodovibrio* or *Thiohalocapsa* spp., rather than to the genera *Lamprocystis* or *Amoebobacter* (Klepac-Ceraj et al., 2012; Hamilton et al., submitted). In addition to photoautotrophy, autotrophic sulfide and sulfur oxidation and extensive sulfate reduction complete the sulfur and carbon cycles in the chemocline (Overmann, 1997; Hamilton et al., submitted). Measurements of exported sulfur—which accumulates intracellularly in *Chromatiaceae*—show that < 0.2% of the upwelling flux of sulfur (as H<sub>2</sub>S and SO<sub>4</sub><sup>2-</sup>) sinks from the chemocline (Overmann et al., 1996a). Instead it has been proposed that this system loses its organic matter through a combination of *in situ* respiration, plus upwards-vertical and lateral rafting of organic debris (Overmann et al., 1994; Overmann et al., 1996b). Lateral rafting brings planktonic material to the shoreline, where it is deposited among a well-developed microbial mat community. The littoral sediments are a mixture of degraded, rafted material, plus organic matter and weathering products from the surrounding landscape and authigenic mineral formations (visible salt crusts; the major lake ions are Mg<sup>2+</sup>, Ca<sup>2+</sup>, Na<sup>+</sup>, SO<sub>4</sub><sup>2-</sup>, and CO<sub>3</sub><sup>2-</sup>; Northcote and Hall, 1983). The shoreline mats also host their own population of PSB, including the cultured species *Thiorhodovibrio winogradskyi*, which has higher oxygen, salinity, and phototolerance than typical planktonic *Chromatiaceae* spp. (Overmann et al., 1992).

Sedimentary evidence shows that euxinia has predominated in Mahoney Lake for at least 9 Kyr (Lowe et al., 1997; Coolen and Overmann, 1998). Although direct export of the 7 m PSB community to sediments is very low, it is commonly believed that much of the organic matter buried in the deep lake sediments ultimately comes from the biomass-rich chemocline. Evidence includes the presence in sediments of the carotenoid biomarker, okenone (Coolen and Overmann, 1998). This compound is found definitively only in planktonic species of PSB (Brocks and Schaeffer, 2008), although there now is debate about its potential origins from microbial mat sources as well (e.g., Meyer et al., 2011). It remains unclear how (or if) the planktonic PSB layer of Mahoney Lake physically reaches the lake-bottom sediments to be preserved. PSB primary production generally doesn't enter macrobiotic ecosystems (Fry, 1986); and in addition, Mahoney Lake is eukaryote-poor, including being free of fish (Northcote and Halsey, 1969).

Because of these many unusual characteristics, Mahoney Lake has been proposed as a modern analogue for early Earth systems that may have been sulfidic into the shallow photic zone (Canfield and Teske, 1996; Brocks et al., 2005; Lyons and Reinhard, 2009; Meyer and Kump, 2008), including environments in which sulfidic photic zones intercept sulfide-oxidizing shallow mats (Meyer et al., 2011; Voorhies et al., 2012). Here we use concentration profiles and compound-specific  $\delta^{13}\text{C}$  analyses of lipids to attempt to resolve the sources and processes governing organic matter burial in Mahoney Lake.

## **Methods**

Water samples from 7 m and 8 m depth in Mahoney Lake (49°17'N, 119°35'W), a grab core of

the underlying lake sediment (15 m depth), and a sample of shoreline microbial mat were collected in July 2008 and stored on dry ice (-70°C). Upon return to lab, water samples were centrifuged in 50 ml Falcon tubes to pellet biomass; all samples subsequently were stored at -80°C until use. Glass and metalware used in sample handling was combusted (450°C, 6 hr); and all Teflon caps, stopcocks, and syringes were rinsed with methanol (MeOH), dichloromethane (DCM), and hexane. Water was Nanopure® grade, and solvents were Burdick & Jackson GC<sup>2</sup> or equivalent.

Total lipid extracts (TLEs) of each sample were obtained using a modified Bligh and Dyer extraction (Bligh and Dyer, 1959; Nishihara and Koga, 1987). Five g of shoreline and lake-bottom sediment or 0.24 g of 7 m and 8 m biomass were extracted in 1:2:0.8 chloroform:MeOH:H<sub>2</sub>O, where the H<sub>2</sub>O contained 0.5% trichloroacetic acid. Each sample was sonicated and vortexed (3 x 5 min) with 5 min on ice (0°C) between each round. Phase separation was achieved by changing the solvent ratio to 1:1:0.8. The organic layer was removed and the aqueous layer was extracted again with chloroform. The combined extracts were washed against 1M NaCl once and H<sub>2</sub>O twice. The resulting TLEs were dried under high-purity nitrogen gas and Na<sub>2</sub>SO<sub>4</sub> was added to remove residual water. TLEs (in DCM) were desulfurized in columns containing 2g Na<sub>2</sub>SO<sub>4</sub> plus 10 g HCl-activated copper pellets. Each TLE was quantitatively transferred to a 2 ml vial, dried just to completeness, and weighed.

TLEs were separated into polarity fractions on columns containing 15 ml of silica gel (70-230 mesh). Fractions were eluted with two column volumes of: hexane (A), 5% ethyl acetate in hexane (B), 15% ethyl acetate in hexane (C), 20% ethyl acetate in hexane (D), 25% ethyl acetate



in hexane (E), 75% ethyl acetate and 25% methanol (F), and methanol (G). Fractions were stored at -20°C until they were analyzed.

For analysis of sterols and *n*-alcohols, trimethylsilyl (TMS) derivatives were prepared from 5% of fractions C, D, and E from the sediment and shoreline and 20% of fractions C, D, and E from the 7 m and 8m using equal parts pyridine and bis(trimethylsilyl)trifluoroacetamide (BSTFA) containing 1% trimethylchlorosilane (TMCS). These derivatives were analyzed relative to injections of a quantitative external standard (stigmasterol) on an Agilent 6890N gas chromatograph with a 30 m x 0.25 mm x 0.25  $\mu$ m Agilent DB-5MS column coupled to an Agilent 5973 mass spectrometer (GC-MS). The PTV inlet of the instrument was maintained at 112°C for 0.85 min, then ramped to 320°C (720°C/min), held for 2.35 min, and ramped to 450°C (720°C/min; hold 5 min) before returning to 112°C. Simultaneously, the oven was maintained at 112°C for 2 min, ramped to 130°C (20°C/min), then to 280°C (6°C/min), then 320°C (3°C/min; hold 25 min). Helium carrier gas flow rate was 1.2 ml/min, the MS transfer line was held at 300°C, and the mass spectrometer scanned  $m/z$  50-750 at 70 eV.

For analysis of fatty acids derived from intact polar lipids (IPLs), 50% of fractions F and G were transesterified with known- $\delta^{13}\text{C}$  methanol containing 5% HCl (4 hr, 70°C) (White et al., 1979; Pearson et al., 2001). Fatty acid methyl esters (FAMES) were extracted using 9:1 hexane:DCM, dried using  $\text{Na}_2\text{SO}_4$ , and analyzed by GC-MS. The GC-MS program was identical to the TMS-derivative program, but with inlet and oven starting temperatures of 65°C and a carrier gas flow rate of 1.4 ml/min. External standards of nonadecanoic acid FAME ( $\text{C}_{19:0}$ -FAME) were used to prepare calibration curves for quantitation; the response factor was linear over the range

corresponding to reported sample peaks. All TMS and FAME derivatives were identified, integrated, and quantified using AMDIS (Stein, 1999). Detection limits ( $3 \times$  signal/noise) for FAMES and sterols were 0.36 ng and 2.16 ng, respectively, as calculated from external standards. This translates into FAME detection limits of ca. 0.1 ng/mg TLE for the sediment and 7 m samples, 0.5 ng/mg TLE for the 8 m sample, and 24 pg/mg TLE for the shoreline sample.

For analysis of  $\delta^{13}\text{C}$  values, FAMES were run on a Thermo Trace Ultra GC (60 m x 0.32 mm x 0.25  $\mu\text{m}$  Agilent DB5-MS column) with a GC Isolink combustion interface coupled to a Thermo Delta V isotope ratio mass spectrometer (GC-C-IRMS). The injector was operated in splitless mode at 220°C. The oven program was 60°C, then 10°C/min to 145°C, 2.7°C/min to 230°C, 5°C/min to 275°C, and 7°C/min to 310°C (5 min hold). Helium carrier gas flow rate was 1.2 ml/min and compounds were combusted at 980°C to  $\text{CO}_2$ . An alkane ( $n\text{-C}_{32}$ ) of known  $\delta^{13}\text{C}$  value (<http://geology.indiana.edu/schimmelmann/>) was co-injected as an internal standard, and each chromatogram was bracketed by  $\text{CO}_2$  reference gas pulses. Isotope chromatograms were integrated in Isodat 3.0 using default background settings, max peak width of 180 sec, and peak resolution of 50%. Values of  $\delta^{13}\text{C}$  from the  $\text{CO}_2$  reference gas were used to correct all compound-specific  $\delta^{13}\text{C}$  values for instrumental drift, assuming a linear trend with time. Peaks with amplitudes  $< 0.5$  V or  $> 10$  V ( $m/z$  44) were eliminated due to non-linear  $m/z$  44 responses. Remaining minor, but statistically significant, size-dependent biasing effects on  $\delta^{13}\text{C}$  values were removed using a linear model in which dilution-series data from external FAME standards of  $\text{C}_{16:0}$ ,  $\text{C}_{19:0}$  and  $\text{C}_{24:0}$  run on the same day were fitted using least-squares approaches (calculated using the *lm* and *predict* functions in the *R* statistical package). Finally, isotope mass balance was used to correct for the  $\delta^{13}\text{C}$  contribution of the FAME derivative methyl carbon.

Bulk values of  $\delta^{13}\text{C}$  were obtained on acidified (10% HCl, 4 hr, 60°C) aliquots of biomass or sediment prepared in tin capsules (Costech 5 x 9 mm). Measurements were made at the MBL Stable Isotope Lab (<http://dryas.mbl.edu/silab/>) and reported relative to the VPDB standard.

## **Results**

### *Fatty Acids*

Fatty acids from fractions F and G displayed similar patterns of abundance relative to each other within a given sample (Figure 2.1a). The higher concentration yield of the F (nominally "glycolipid", but also including sulfolipids and other less polar IPL head-groups; Schubotz et al., 2009; Close et al., 2013) fraction relative to the G (nominally "phospholipid") fraction may be due to partial loss of charged phospholipid compounds during extraction and/or separation on silica gel (Huguet et al., 2010). However, sufficient yield was obtained from all G samples for quantitative analysis and for measurement of  $\delta^{13}\text{C}$  values, with the exception of the shoreline G sample (Figure 2.1a,b).

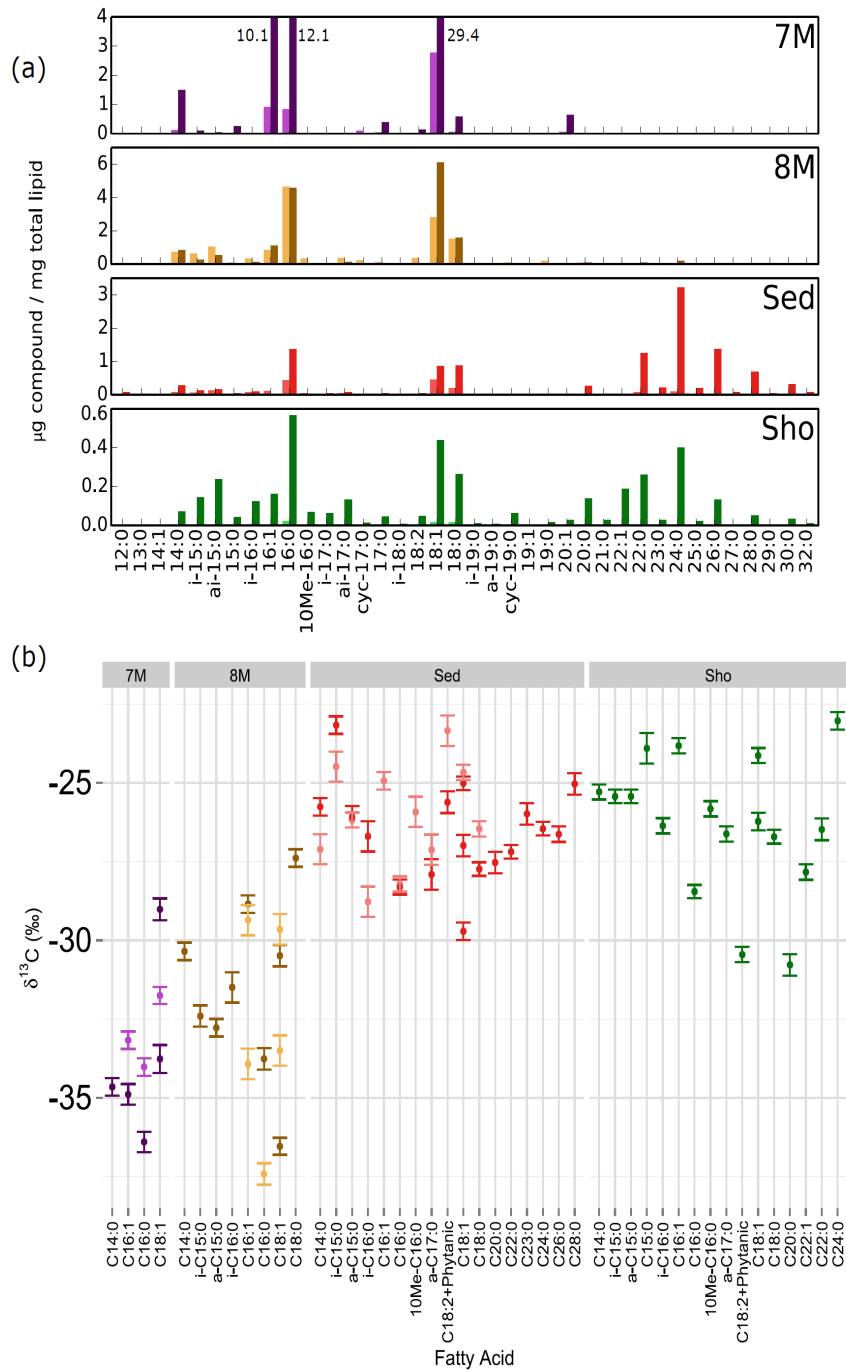


Figure 2.1: (a) Fatty acid concentrations as a percentage of total lipid extract, arranged by sample. Fractions F (lighter colors) and G (darker colors) are plotted together. (b) Carbon isotope values of individual fatty acids sorted by sample from fractions F (lighter colors) and G (darker colors).

The samples qualitatively fall into two groups: water-column organic matter vs. benthic organic matter. The 7 m PSB sample from the chemocline primarily contains  $C_{18:1}$ , lesser amounts of  $C_{16:0}$  and  $C_{16:1}$ , and minor amounts of  $C_{14:0}$ ,  $C_{18:0}$ , and  $C_{20:1}$ . Both the relative and absolute quantities of other trace fatty acids in this sample are below those of the other three samples (Figure 2.1a; Appendix 1, Table S1). The ratio of total  $C_{18}$  chain lengths ( $C_{18:0} + C_{18:1}$ ) to  $C_{16}$  chain lengths ( $C_{16:0} + C_{16:1}$ ) is 1.35 (calculated for fraction F). There are no fatty acids of chain length  $> C_{20}$  that are above detection limits. The 8 m sample represents material from the dark hypolimnion which was sampled well below the chemocline, yet its fatty acid composition is broadly similar to the chemocline:  $C_{18:1}$  is the dominant compound, but with relatively more  $C_{16:0}$  and  $C_{18:0}$ . Here the ratio of  $C_{18}$  chain lengths ( $C_{18:0} + C_{18:1} + C_{18:2}$ ) to  $C_{16}$  chain lengths ( $C_{16:0} + C_{16:1} + iso-C_{16:0}$ ) is 1.32 (fraction F). At 8 m there also are detectable fatty acids of chain length  $> C_{20}$ , including minor amounts of  $C_{22:0}$ ,  $C_{24:0}$ , and  $C_{26:0}$  (maximum at  $C_{24:0}$ ). The ratio of  $C_{24:0}$  to the sum of  $C_{16}$  chain lengths ( $C_{16:0} + C_{16:1} + iso-C_{16:0}$ ), however, is only 0.04, indicating that both the 7 m and 8 m samples primarily consist of  $C_{16}$  and  $C_{18}$  fatty acids.

The shoreline and sediment samples quantitatively and qualitatively have a greater diversity of fatty acids (Figure 2.1A; Figure 2.2), with both having almost twice as many AMDIS detectable components relative to the 7 m and 8 m samples. The sediment sample contains a regular series of longer-chain fatty acids with a strong even-over-odd predominance and a maximum at  $C_{24:0}$ . In the sediment,  $C_{18:1}$ ,  $C_{18:0}$ , and  $C_{16:0}$  fatty acids are approximately equally abundant, and the ratio of total  $C_{18}$  chain lengths to  $C_{16}$  chain lengths is 1.21. The longer-chain components dominate this sample and the ratio of  $C_{24:0}$  to the sum of  $C_{16}$  chain lengths is 2.16. The shoreline sample

resembles the basinal sediment. It contains the regular series of longer-chain fatty acids, also with a maximum at C<sub>24:0</sub>. C<sub>16:0</sub> is the dominant shorter-chain compound, although C<sub>18:1</sub> and C<sub>18:0</sub> are also major components. The ratio of total C<sub>18</sub> chain lengths to C<sub>16</sub> chain lengths is 0.89, and the ratio of C<sub>24:0</sub> to the sum of C<sub>16</sub> chain lengths is 0.47.

Specific bacterial fatty acids are widespread in all samples. The lipid 10-methyl-C<sub>16:0</sub> is characteristic of anoxic sulfur cycling, being a common component of sulfate reducers such as *Desulfobacter* spp. (Taylor and Parkes, 1983). It was present in all samples except the 7 m sample, and its abundance was highest in the shoreline (Figure 2.1a; Appendix 1, Table S1). All samples also contained detectable quantities of the bacterial lipids *iso*- and *anteiso*-C<sub>15:0</sub> (Kaneda, 1991), and most contained *iso*- and *anteiso*-C<sub>17:0</sub> as well. These components were particularly abundant in the shoreline sample (ratio of branched C<sub>15</sub> compounds to C<sub>16:0</sub> of 0.67) with lesser amounts in the 8m and sediment samples (branched C<sub>15</sub> to C<sub>16:0</sub> ratios of 0.18 and 0.22, respectively). Polyunsaturated fatty acids were not detected in any sample.

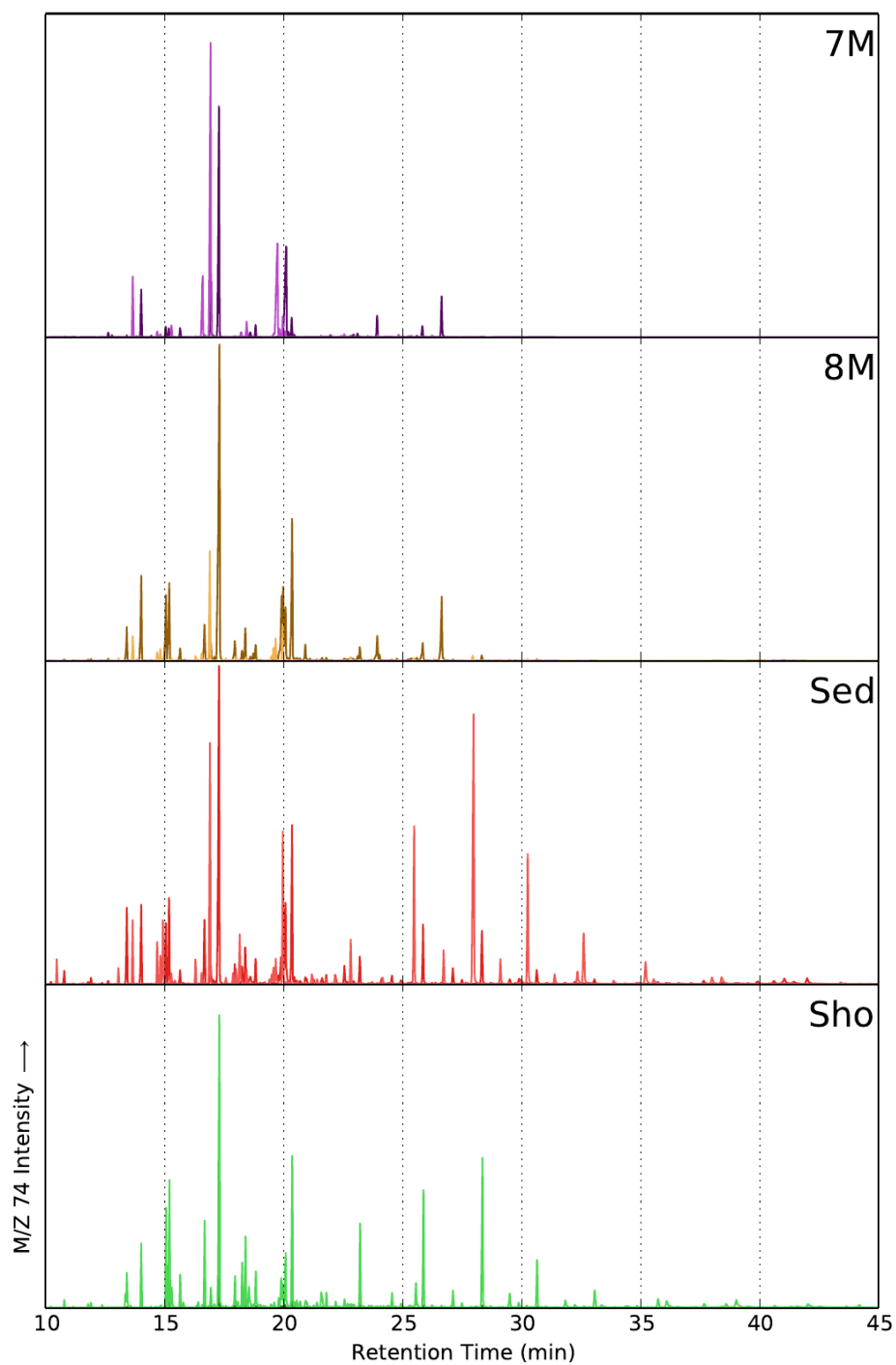


Figure 2.2: Extracted Ion chromatograms (EICs) at  $m/z$  74 of GC-MS runs of fractions F (lighter colors) and G (darker colors) from each sample. The GC program is described in the main text.

### *Fatty Acid Isotopes*

Fatty acid  $\delta^{13}\text{C}$  values (Figure 2.1b) also distinguish the shoreline and sediment samples as one group, and the 7-m and 8-m samples as a separate group. Values for individual compounds from the water column are relatively depleted in  $^{13}\text{C}$ , ranging from -31.4‰ to -36.4‰ at 7 m and from -27.4‰ to -37.4‰ at 8 m. These numbers are similar to other measurements of purple sulfur bacterial populations in meromictic lakes (Hartgers et al., 2000). The low diversity of the 7 m sample yielded only four fatty acids above the analytical threshold for GC-C-IRMS analysis. In contrast, more compounds could be analyzed at 8 m. However, when comparing all of the 7 m and 8 m data there appear to be no systematic patterns, either inter- or intra- sample, or among the fatty acids. For example, the odd-chain and branched lipids plot within the range of the non-branched components, and unsaturated lipids are neither systematically enriched nor depleted in  $^{13}\text{C}$ . The mass-weighted average  $\delta^{13}\text{C}$  value for all fatty acids at 7 m is -33.3‰, while at 8 m it is -32.4‰.

In comparison, fatty acid values of  $\delta^{13}\text{C}$  from the shoreline and sediment samples are enriched in  $^{13}\text{C}$ , with a mass-weighted average of -26.2‰ for the shoreline and -26.8 for the sediment (range -23.0‰ to -30.8‰ for shoreline, and -23.2‰ to -28.8‰ for sediment). Among saturated compounds, there is minimal isotopic difference between samples. For example,  $\text{C}_{16:0}$  measures -28.3‰ and -28.4‰ in the sediment and shoreline, respectively (fraction F). The heterogeneity is larger for unsaturated and branched compounds, but again without notable patterns.

Unusually, absence of isotopic patterns extends also to the long-chain compounds,  $\text{C}_{22:0}$ - $\text{C}_{28:0}$ . We do not see evidence for a sawtooth pattern of  $^{13}\text{C}$  enrichment in even-chain lengths relative to



odd-chain lengths. We also do not see a systematic decrease in overall  $^{13}\text{C}$  content of the long-chain compounds relative to the short-chain compounds (Drenzek et al., 2007; Gong and Hollander, 1997; Naraoka & Ishiwatari, 2000), *i.e.*, the  $\delta^{13}\text{C}$  values of shoreline and sedimentary long chain fatty acids are within the range of the shorter chain fatty acids and do not differ from their mean by more than a few mil.

### *Sterols & Fatty Alcohols*

Both the shoreline and the 7 m sample are low in total sterol content, while sterols are a higher proportion of total lipids in the 8 m sample and the sediments (Figure 2.3a). The sterol distribution in all samples is enriched in  $\text{C}_{29}$  sterols (50-90% of total) relative to  $\text{C}_{27}$  sterols (10-37%), with low amounts of  $\text{C}_{28}$  sterols ( $\leq 12\%$ ).

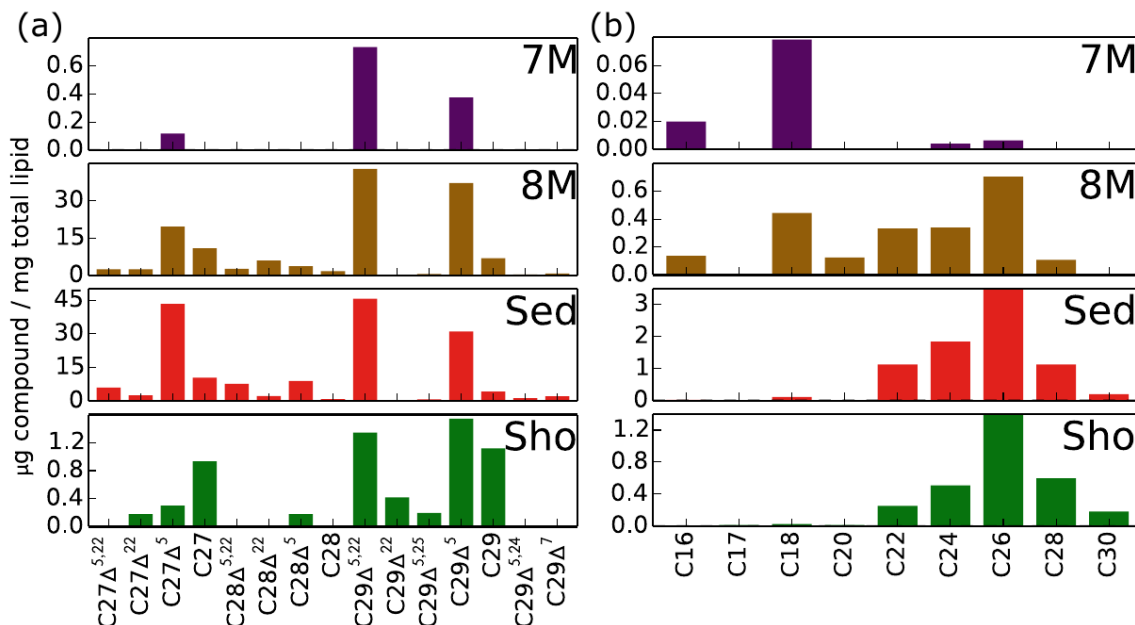


Figure 2.3: (a) Sterol concentrations as a percentage of total lipid extract. (b) n-alcohol concentrations as a percentage of total lipid extract.

Concentrations of *n*-alcohols are low in all samples. The primary fatty alcohol at 7 m is *n*-C<sub>18</sub> with a lesser amount of *n*-C<sub>16</sub> (Figure 2.3b). Elevated levels of C<sub>18</sub> and C<sub>16</sub> fatty alcohols also are commonly observed in marine systems (Wakeham, 1982; Mudge and Norris, 1997). The shoreline contains a continuous series of *n*-alcohols with strong even-over odd predominance, peaking at C<sub>26</sub>. The *n*-alcohol profile from the sediment sample also strongly resembles the shoreline profile. Uniquely, and unlike for the fatty acid profiles, the *n*-alcohol fraction from 8 m appears to resemble a mixture between the composition at 7 m and the composition in the shoreline.

*Table 2.1:  $\delta^{13}\text{C}$  values for Mahoney Lake bulk TOC.*

<b>Sample</b>	<b>Bulk <math>\delta^{13}\text{C}</math> (‰)</b>	<b>Weight % C</b>
Shoreline	-23.6, -24.0	3.4%
7 m	-27.2	n/a
8 m	-27.2	n/a
Sediment	-24.2, -25.0	20.5%

#### *Bulk Carbon*

Bulk organic carbon isotopes approximately track the fatty acid  $\delta^{13}\text{C}$  values, with the samples again falling in two groups:  $^{13}\text{C}$ -enriched sediments and  $^{13}\text{C}$ -poor planktonic material. In the water column, total organic carbon (TOC) values (-27.2‰ for both samples; Table 2.1) are 6.6‰ and 4.7‰ enriched in  $^{13}\text{C}$  relative to average fatty acids at 7 m and 8 m, respectively. The offsets

for the sediment and shoreline samples are smaller: sediment TOC (average -24.6‰) is 1.7‰ enriched in  $^{13}\text{C}$  compared to fatty acids, while the shoreline TOC is 2.5‰ enriched in  $^{13}\text{C}$  compared to fatty acids (Table 2.1).

## Discussion

### *Compound distributions*

The purple sulfur bacteria, *Chromatiaceae*, typically have fatty acid patterns with ca. 25%  $\text{C}_{16:0}$ , 30%  $\text{C}_{16:1}$ , 40%  $\text{C}_{18:1}$ , and minor amounts of  $\text{C}_{14:0}$  and  $\text{C}_{18:0}$  (Imhoff and Bias-Imhoff, 1995). The profile of the 7 m sample would be consistent with almost all of the fatty acids coming from these phototrophs, with the exception of  $\text{C}_{20:1}$  and a few other trace components. Indeed, biomass in the chemocline of Mahoney Lake is reported to be dominated by a single species, *Chromatiaceae* sp. ML1 (Overmann et al., 1991), which is now classified as a *Thiohalocapsa* sp. (Tank et al., 2009; Hamilton et al., submitted). Cell counts by dilution (Overmann et al., 1991) and metagenomic coverage data (Hamilton et al., submitted) indicate that this species may account for >90% of the biomass in the phototrophic layer at 7 m. Based on the similarity of fatty acid profiles, the material at 8 m also may contain sinking biomass of *Chromatiaceae* sp. ML1, although with a larger fraction of other species that supply such compounds as 10-Me- $\text{C}_{16:0}$  (*i.e.*, sulfate reducers). The sample from 8 m also contains higher relative abundances of the minor fatty acids, again pointing to an additional source that has different endmember composition.

Clues to this source are found by examining the alcohol fractions. The shoreline and sediment

samples are notable for having abundant long-chain saturated fatty acids and alcohols, with regular distribution around C<sub>24:0</sub>. Such patterns are commonly explained by a strong contribution from terrestrial plants (Cranwell, 1981; Rao et al., 2009). Low levels of C<sub>28</sub> sterols and a predominance of C<sub>29</sub> over C<sub>27</sub> sterols also is typical of terrestrial plant inputs (Huang and Meinshein, 1979; Volkman, 2003). Cawker (1983) found that 60% of the pollen preserved in recent Mahoney Lake sediment is from *Pinus* (Pine), with slightly < 10% each of *Pseudotsuga* (Fir), *Alnus* (Alder), and *Poaceae* (grasses). The surrounding vegetation also is believed to contribute to the high concentration of dissolved organic carbon (DOC) in the lake (Hall and Northcote, 1990; Overmann, 1997).

Looking solely at the distribution of lipid biomarkers in all four samples would suggest that planktonic material mainly is remineralized in the water column, while the sediments derive primarily from terrestrial organic matter, reflecting a combination of eroded shoreline sediments and epilimnic sinking material (possibly of aeolian source). This would be consistent with the carbon and sulfur budget proposed by Overmann (1997). However, it fails to explain the high concentrations of okenone and the presumed quantitative importance of PSB inputs to the basinal sediments (Overmann et al., 1993; Coolen and Overmann, 1998). Similarly, it is difficult to reconcile fully with <sup>13</sup>C distributions.

#### *Constraints from $\delta^{13}\text{C}$ values*

Although the isotope ratios also show the presence of distinct endmembers in this system, they do not necessarily point to a clear planktonic vs. terrigenous distinction. With average values

around -33‰ and -27‰ for individual fatty acids and bulk materials, respectively, the 7 m samples are consistent with a primary source from bacterial autotrophy. *Chromatiaceae* sp. ML1 and other related PSB use Rubisco Type 1A to fix dissolved CO<sub>2</sub> (Badger and Bek, 2008; Hamilton et al., submitted). The chemocline of Mahoney Lake contains ca. 40 mM dissolved inorganic carbon at pH 8.1, suggesting CO<sub>2</sub> should not be excessively limited and that typical fractionations for Type 1A Rubisco ( $\epsilon = 20 - 24\text{‰}$ ; Sirevåg et al., 1977; Madigan et al., 1989; Scott et al., 2007) could generate the values observed. Similar or even more negative values have been seen in other contemporary stratified systems (e.g., Hartgers et al., 2000; Velinsky and Fogel 1999). Our fatty acid and bulk biomass values at 7 m also are broadly consistent with previous reports for purified okenone and bacteriochlorophyll *a* from the Mahoney Lake water column ( $-27.2 \pm 1.9\text{‰}$ ;  $-27.8 \pm 0.7\text{‰}$ ; Overmann et al., 1996b). Slight enrichment of <sup>13</sup>C in fatty acids from 8 m (average value ca. -32‰ rather than -33‰) would be consistent with the presence of a minor *in situ* endmember or with heterotrophic reworking of the material descending from 7 m (DeNiro and Epstein, 1978; Hayes, 1993).

Biomarker abundance profiles from the sediments point to significant allochthonous overprinting of the original planktonic production, with their compositional patterns of *n*-alcohols, long-chain fatty acids, and sterols typical of higher plants. Unfortunately, we could not measure values of  $\delta^{13}\text{C}$  for *n*-alcohols or sterols, precluding definitive knowledge of the terrigenous endmember. The sediment  $\delta^{13}\text{C}$  data for fatty acids (ca. -26‰) also argue against sedimentation of PSB biomass and for dominant admixture of <sup>13</sup>C-enriched material from elsewhere in the lake or its catchment.

However, the interpretation that there is a strong C<sub>3</sub> plant contribution to the basinal sediments is complicated by the compound-specific  $\delta^{13}\text{C}$  data. There is no statistical difference between the observed  $\delta^{13}\text{C}$  values for long-chain and shorter-chain fatty acids in the shoreline or sediment samples ( $< \text{C}_{20}$  chain lengths vs.  $\geq \text{C}_{20}$  chain lengths; two-tailed  $t$ -test), a potential argument that all of these compounds have the same source. By contrast, in sediments of the Mackenzie Shelf, values of  $\delta^{13}\text{C}$  for fatty acids show a stepwise decrease from ca. -26‰ (chain lengths  $< \text{C}_{20}$ ) to ca. -31‰ (chain lengths  $\geq \text{C}_{20}$ ) (Drenzek et al., 2007). A similar pattern also is observed for fatty acids of the Washington Margin (Feng et al., 2013). Such a bimodal distribution, with  $^{13}\text{C}$ -depleted values for long-chain compounds, is interpreted to reflect the influence of C<sub>3</sub>-dominated plant detritus only at longer chain lengths. This pattern also would be expected for Mahoney Lake sediments if the fatty acids reflected mixed sources; alternatively, if the lake sediments contained solely terrigenous lipids, then all compounds might be expected to have values ca. -31‰, not near -26‰.

Values of  $\delta^{13}\text{C}$  measured for pine needles and needle litter in the area are -27.2‰ and -24.5‰, respectively (Overmann et al., 1996b). The latter value is a possible explanation for the  $^{13}\text{C}$ -enriched bulk values observed in shoreline and sediment samples. The total system may be fed by detrital plant matter having an endmember value somewhat atypical ( $^{13}\text{C}$ -enriched) for C<sub>3</sub> environments. However, fatty acids would be derived from the primary pine biomass and should retain the biosynthetic signature, *i.e.*,  $\delta^{13}\text{C}$  values ca. 3-5‰ lower than the non-degraded needles (Hayes, 1993). Again this would predict terrigenous fatty acid values ca. -31‰, which is inconsistent with the values measured for the shoreline and basinal sediment samples.

Bacterial sources occasionally have been proposed as an explanation for some environmental long-chain fatty acid profiles (Volkman et al., 1988; Gong and Hollander 1997). The most striking examples – both for their similarity to the profiles observed here, as well as for their definitive microbial origin – are the profiles of fatty acids extracted from modern ooids (Summons et al., 2013). Like our samples, the ooids have long-chain fatty acids with maxima at C<sub>24</sub>, similar proportions of C<sub>14</sub>-C<sub>18</sub> fatty acids, and abundant branched-chain and 10-Me-C<sub>16:0</sub> structures. Some bacteria produce long-chain compounds in other contexts (*e.g.*, heterocyst glycolipids of Cyanobacteria; Gambacorta et al., 1998), but only recently has it been suggested that long-chain polyketide synthesis pathways may be widespread among environmental bacteria (Shulze and Allen, 2011). Direct synthesis by the microbial mat and/or littoral bacterial community potentially could explain both the Mahoney and the ooid observations.

The abundant short-chain fatty acids in the shoreline, especially those that are methylated or have cyclopropyl moieties, also indicate a significant bacterial contribution to this environment (Perry et al., 1979). Specifically, the combined abundance of bacterial *i*-C<sub>15:0</sub> and *a*-C<sub>15:0</sub> (Kaneda, 1991) relative to *n*-C<sub>15:0</sub> fatty acids (ratios of 9.3 and 8.9 in the shoreline and sediment samples, respectively) is even higher than typically reported for bacterially-dominated sediments (4 to 4.5; Parkes and Taylor, 1983). This suggests that despite its proximity to surrounding vegetation, the organic matter in the Mahoney Lake shoreline actually is dominated by the rich microbial mat community (Overmann et al., 1992). This community may be sustained by a combination of degraded higher plant detritus and rafting PSB material, which is then recycled and re-synthesized, yielding *de novo* fatty acids enriched in <sup>13</sup>C.

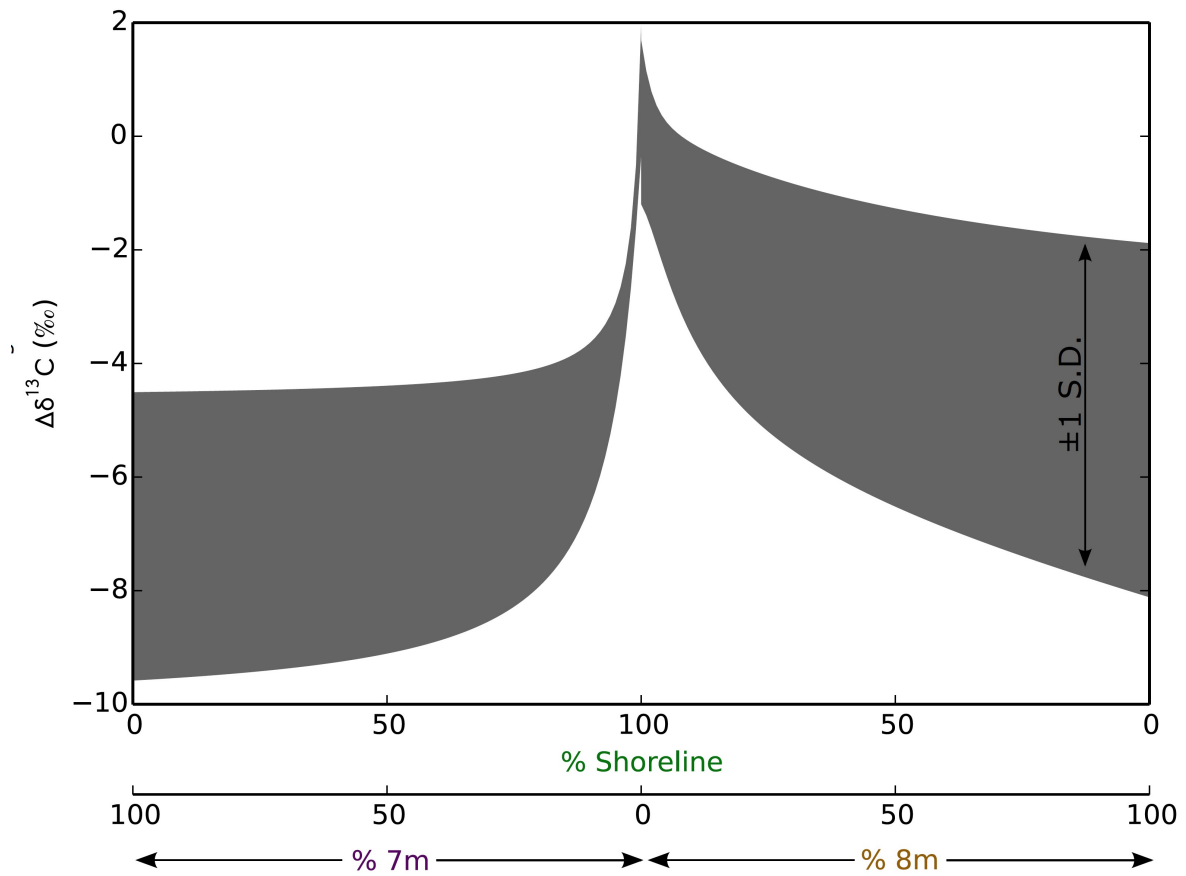


Figure 2.4: (a) Mixing model of the predicted offset between fatty acid  $\delta^{13}\text{C}$  values of sediment and a concentration-weighted mixture of the shoreline sample with either 7-meter (left) or 8-meter (right) samples. Values closer to zero indicate a better fit of the modeled mixture to the true sedimentary lipid  $\delta^{13}\text{C}$  values.

#### Mixing model of sources to lake-bottom sediment

We constructed an isotope mass-balance mixing model to determine how much autochthonous organic material (from 7 m or 8 m) could be contributing to the basinal sediments while still being concealed by the primarily allochthonous (shoreline-derived) inputs. Our mixing model is modified from Collister et al. (1994) and uses fatty acid concentrations and  $\delta^{13}\text{C}$  values to



determine the fractional contributions of two endmembers to a total mixture. For each compound we computed the predicted sediment value ( $\delta_{sed}$ ) over the full range ( $x = 0$  to  $1$ ) of mixing the autochthonous component ( $x$ ) with shoreline-derived organic matter:

$$(1) \quad \Delta\delta_i = \frac{xM_{auto,i}\delta_{auto,i} + (1-x)M_{shore,i}\delta_{shore,i}}{xM_{auto,i} + (1-x)M_{shore,i}} - \delta_{sed,i}$$

Here  $i$  is an individual compound,  $\delta_i$  is the  $\delta^{13}\text{C}$  value of  $i$  in a given endmember, and  $M_i$  is its concentration. This yields the difference ( $\Delta\delta_i$ ) between the observed  $\delta_i$  value for a given compound in the sediment ( $\delta_{sed,i}$ ) and the value predicted for that compound if its source is a mixture of the two endmembers, *i.e.*, the best fit is  $\Delta\delta_i = 0\text{‰}$ . The mean of these differences ( $\Delta\delta \pm 1\sigma$ ; averaged over all  $\Delta\delta_i$ ) is shown as the filled region in Figure 2.4. The two mixing scenarios are the 7 m PSB layer mixed with shoreline (left side of figure) and the 8 m sample mixed with shoreline (right side of figure). We used the 8 m ("sulfidic zone") sample as one of the two autochthonous endmembers on the assumption that it integrates contributions from all of the lake's autotrophic producers, plus aeolian delivery of terrigenous plant inputs. This was necessary because the upper photic zone (oxic mixolimnion shallower than 7 m) is oligotrophic and has minimal primary production from cyanobacteria and Chrysophyta (Northcote and Hall, 1983) – thus we did not have sufficient shallow-water biomass for lipid analysis, despite having previously analyzed DNA from 5 m (Klepac-Ceraj et al., 2012). In support of this strategy, genomic data show the 8 m sample contains numerous sequences linked to eukaryotes incapable of living in sulfidic waters (Hamilton et al, submitted), consistent with a contribution of organic matter from sinking algal production and/or other plants.

Regardless, the model shows that the sediment cannot contain much of either planktonic endmember, although it does show the sediments are able to “conceal” some material having the composition of the 8-m endmember. The maximum amount of 7-m material that could be added while staying within the one sigma envelope is 1%, while the maximum amount of 8-m material is 46% (to satisfy  $\Delta\delta^{13}\text{C} = 0 \pm 1\text{‰}$ ). It is likely that the 8 m endmember already contains a shoreline contribution, however, making the latter number effectively smaller. More importantly, while we consider it likely that some of the biomass of PSB from 7 m ultimately makes it to the basinal sediments, the model provides evidence that the route first involves rafting to the shoreline where it must be heavily processed.

Our conclusion that the majority of the sediment flux comes from the shoreline is unexpected, but it is not necessarily incompatible with the literature. Coolen and Overmann (1998) found okenone in the deep basin sediments, but the abundance of preserved *Chromatiaceae* sp. ML1 DNA was low and was not correlated with okenone concentrations. Given our data, it is possible—although far from demonstrated—that the sediment okenone derives from a PSB community that lives in the shoreline and is taxonomically different from the 7-m community. This would support a possible benthic or mat-derived source for okenone (*e.g.*, Meyer et al., 2011) and is contrary to the prevailing hypothesis that okenone is necessarily planktonic (Brocks and Schaeffer, 2008). It is notable that the -27.2‰ value reported for okenone from basinal sediments (Overmann et al., 1996b) is more similar to our shoreline and sediment fatty acids than to fatty acids from the water column. The typical  $^{13}\text{C}$  enrichment in isoprenoid relative to acetogenic lipids (ca. 0-3‰; Hayes, 2001) is too small to allow the -27‰ value for okenone to have the

same source as the water column fatty acids (ca. -33‰); the okenone either has been affected by diagenetic fractionation or it originates *de novo* in the shoreline. In support of the diagenetic explanation, the only PSB taxon characterized to date from the shoreline does not contain okenone (Overmann et al., 1992). Yet a similar diagenetic argument cannot be made to explain the fatty acid data, because as discussed previously, the abundant long-chain *n*-alkyl compounds do not derive from the water column – they must be added in the shoreline. A possible means to reconcile all observations is to invoke rafting of PSB debris to the shoreline, where it is mixed with terrigenous plant detritus. Intensive bacterial heterotrophy within the microbial mat system would resynthesize all *n*-alkyl lipids (Logan et al., 1995; Close et al., 2011), and this may be concomitant with significant <sup>13</sup>C fractionation of okenone.

#### *Processes Controlling Delivery of Organic Matter to Basinal Sediments*

The failure of the PSB layer to sink to the sediments has been noted previously. Buoyant rafting is associated with seasonal die-off, and estimates suggest >85% of the PSB are lost by this route (Overmann et al., 1994). There also is a relatively small flux of elemental sulfur (S<sup>0</sup>) and biomass to sediment traps below the chemocline (Overmann et al., 1996a,b). Nearly all of the remainder of the PSB and associated community is presumed to be degraded *in situ*. The strong density gradient in the chemocline helps retain this microbial organic matter in neutrally buoyant layers where it may be remineralized in the water column, a feature also common in marine systems (MacIntyre et al., 1995; Sorokin, 2002). Temporal and physical redox oscillations also can occur at these interfaces, significantly increasing the remineralization of organic matter (Aller, 1998). These cycles occur in Mahoney Lake both diurnally and in association with a small seasonal

migration of the chemocline (Hall and Northcote, 1990).

In Mahoney Lake, shoreline-derived organic matter may bypass the intensive remineralization associated with such density stratification due to its association with denser particles. In many marine systems, lateral and down-slope transport is a major contributor of organic matter to sediments (Mollenhauer and Eglinton, 2007; Inthorn et al., 2006). Turbidity currents also have been known to contribute to sedimentary deposition in other meromictic lakes, such as Fayetteville Green Lake (Ludlam, 1974), where they are responsible for ca. 50% of deposition. In such cases, dense inorganic minerals and debris flows may help buoyant organic matter be transported across salinity gradients and into sub-chemocline basins.

Downslope transport of anoxygenic photoautotrophic mats may have been a significant process earlier in Earth's history. Mats of non-photosynthetic sulfide oxidizers such as *Thioploca* cover thousands of square kilometers of the present-day ocean floor where oxygen minimum zones intersect sediments (Otte et al., 1999). In “anoxic” oceans, these sulfide-oxidizing microbial mats may have migrated to shallower depths, intersecting the photic zone and containing PSB. Such sulfide-oxidizing mats today exist under suboxic conditions of Lake Superior sinkholes (Voorhies et al., 2012) and on the margins of Fayetteville Green Lake (Meyer et al., 2011). In such a scenario, the water column may only be suboxic but not euxinic, and the traditional biomarkers (okenone, chlorobactene) for photic zone euxinia (PZE) would be derived from benthic mats (e.g., Meyer et al., 2011). Much as in Mahoney Lake, these biomarkers might then be laterally transported downslope into areas not overlain by PZE, thereby complicating interpretations of planktonic vs. benthic PZE and the level of sulfide in past water columns. This

is a significant issue of mat vs. planktonic biomarker geochemistry that remains to be resolved by future studies.

Several episodes of Earth's history may have such conflicting records. Rock records of the Permo-Triassic boundary contain isorenieratene and aryl isoprenoids (Grice et al., 2005; Hays et al., 2007), but atmospheric oxygen levels may not have been notably lower than today (Knoll et al., 2007) and euxinia may not have necessarily been linked to either paleo-oxygen levels or to the magnitude of the extinction event (Nielsen et al., 2010). Similarly, Neoproterozoic marine rocks contain okenane, chlorobactene, and other diagenetic products of anoxygenic photoautotrophs (Brocks et al., 2005), but iron speciation data suggest the ocean may not have been strongly euxinic (Johnston et al., 2010; Planavsky et al., 2011). Our results suggest that biomarker interpretations of photic zone euxinia should be evaluated in context of local stratigraphy, and lateral transport processes should be considered.

## **Conclusion**

Fatty acid, sterol, and *n*-alcohol concentrations, along with fatty acid  $\delta^{13}\text{C}$  values, show that the majority of organic matter buried in Mahoney Lake comes from shoreline microbial mat production as opposed to water column production. There also is additional, but difficult to quantify, material derived from terrigenous inputs. These findings imply that in stratified systems, particle density and lateral transport are significant agents aiding the process of organic burial. This potentially is an important consideration for interpreting stratigraphic and biomarker evidence for photic zone euxinia.

## References

- Aller RC (1998) Mobile deltaic and continental shelf muds as suboxic, fluidized bed reactors. *Marine Chemistry*, **61**(3-4), 143–155.
- Badger MR, Bek EJ (2008) Multiple Rubisco forms in proteobacteria: their functional significance in relation to CO<sub>2</sub> acquisition by the CBB cycle. *Journal of Experimental Botany*, **59**(7), 1525.
- Bligh EG, Dyer WJ (1959) A rapid method of total lipid extraction and purification. *Can J Physiol Pharmacol*, **37**, 911–917.
- Brocks JJ, Love GD, Summons RE, Knoll AH, Logan GA, Bowden SA (2005) Biomarker evidence for green and purple sulphur bacteria in a stratified Palaeoproterozoic sea. *Nature*, **437**(7060), 866–870.
- Brocks JJ, Schaeffer P (2008) Okenane, a biomarker for purple sulfur bacteria (Chromatiaceae), and other new carotenoid derivatives from the 1640 Ma Barney Creek Formation. *Geochimica et Cosmochimica Acta*, **72**(5), 1396–1414.
- Canfield DE, Teske A (1996) Late Proterozoic rise in atmospheric oxygen concentration inferred from phylogenetic and sulphur-isotope studies. *Nature*, **382**, 127–132.
- Cawker KB (1983) Fire history and grassland vegetation change: three pollen diagrams from southern British Columbia. *Canadian Journal of Botany*, **61**(4), 1126–1139.
- Close HG, Bovee R, Pearson A (2011) Inverse carbon isotope patterns of lipids and kerogen record heterogeneous primary biomass. *Geobiology*, **9**(3), 250–265.
- Close HG, Shah SR, Ingalls AE, Diefendorf AF, Brodie EL, Hansman RL, Freeman KH, Aluwihare LI, Pearson A (2013) Export of submicron particulate organic matter to mesopelagic depth in an oligotrophic gyre. *Proceedings of the National Academy of Sciences*, **110**(31), 12565–12570.
- Collister JW, Lichtfouse E, Hieshima G, Hayes JM (1994) Partial resolution of sources of n-alkanes in the saline portion of the Parachute Creek Member, Green River Formation (Piceance Creek Basin, Colorado). *Organic Geochemistry*, **21**(6–7), 645–659.
- Coolen MJL, Overmann J (1998) Analysis of Subfossil Molecular Remains of Purple Sulfur Bacteria in a Lake Sediment. *Appl. Environ. Microbiol.*, **64**(11), 4513–4521.
- Cranwell PA (1981) Diagenesis of free and bound lipids in terrestrial detritus deposited in a lacustrine sediment. *Organic Geochemistry*, **3**(3), 79–89.

- DeNiro MJ, Epstein S (1978) Influence of diet on the distribution of carbon isotopes in animals. *Geochimica et Cosmochimica Acta*, **42**(5), 495–506.
- Drenzek NJ, Montluçon DB, Yunker MB, Macdonald RW, Eglinton TI (2007) Constraints on the origin of sedimentary organic carbon in the Beaufort Sea from coupled molecular  $^{13}\text{C}$  and  $^{14}\text{C}$  measurements. *Marine Chemistry*, **103**(1–2), 146–162.
- Feng X, Benitez-Nelson BC, Montluçon DB, Prahl FG, McNichol AP, Xu L, Repeta DJ, Eglinton TI (2013)  $^{14}\text{C}$  and  $^{13}\text{C}$  characteristics of higher plant biomarkers in Washington margin surface sediments. *Geochimica et Cosmochimica Acta*, **105**, 14–30.
- Fry B (1986) Sources of carbon and sulfur nutrition for consumers in three meromictic lakes of New York State. *Limnology and oceanography*, **31**(1), 79–88.
- Gambacorta A, Pagnotta E, Romano I, Sodano G, Trincone A (1998) Heterocyst glycolipids from nitrogen-fixing cyanobacteria other than nostocaceae. *Phytochemistry*, **48**(5), 801–805.
- Gong C, Hollander DJ (1997) Differential contribution of bacteria to sedimentary organic matter in oxic and anoxic environments, Santa Monica Basin, California. *Organic Geochemistry*, **26**(9–10), 545–563.
- Grice K, Cao C, Love GD, Twitchett RJ, Bottcher M, Grosjean E, Summons RE, Turgeon SC, Dunning W, Jin Y (2005) Photoc Zone Euxinia During the Permian-Triassic Superanoxic Event. *Science*,
- Hall KJ, Northcote TG (1990) Production and decomposition processes in a saline meromictic lake. *Hydrobiologia*, **197**(1), 115–128.
- Hamilton TL, Bovee RJ, Sattin SR, Mohr W, Schaperdorth I, Gilhooly WP, Lyons TW, Pearson A, Macalady JL (2013) Sulfur dynamo in the phototrophic plate of a meromictic lake. *Geobiology*, (submitted).
- Hartgers WA, Schouten S, Lopez JF, Sinninghe Damsté JS, Grimalt JO (2000)  $^{13}\text{C}$ -contents of sedimentary bacterial lipids in a shallow sulfidic monomictic lake (Lake Cisó, Spain). *Organic Geochemistry*, **31**(9), 777–786.
- Hayes JM (1993) Factors controlling  $^{13}\text{C}$  contents of sedimentary organic compounds: Principles and evidence. *Marine Geology*, **113**(1), 111–125.
- Hayes JM (2001) Fractionation of Carbon and Hydrogen Isotopes in Biosynthetic Processes. *Reviews in Mineralogy and Geochemistry*, **43**(1), 225.
- Hays LE, Beatty T, Henderson CM, Love GD, Summons RE (2007) Evidence for photic zone euxinia through the end-Permian mass extinction in the Panthalassic Ocean (Peace River

- Basin, Western Canada). *Palaeoworld*, **16**(1), 39–50.
- Huang W-Y, Meinschein WG (1979) Sterols as ecological indicators. *Geochimica et Cosmochimica Acta*, **43**(5), 739–745.
- Huguet C, Urakawa H, Martens-Habbena W, Truxal L, Stahl DA, Ingalls AE (2010) Changes in intact membrane lipid content of archaeal cells as an indication of metabolic status. *Organic Geochemistry*, **41**(9), 930–934.
- Imhoff J, Bias-Imhoff U (1995) Lipids, quinones and fatty acids of anoxygenic phototrophic bacteria. In: *Anoxygenic Photosynthetic Bacteria* (eds. Blankenship R, Madigan M, Bauer C). Springer, Dordrecht, pp. 179–205.
- Inthorn M, Wagner T, Scheeder G, Zabel M (2006) Lateral transport controls distribution, quality, and burial of organic matter along continental slopes in high-productivity areas. *Geology*, **34**(3), 205–208.
- Johnston DT, Poulton SW, Dehler C, Porter S, Husson J, Canfield DE, Knoll AH (2010) An emerging picture of Neoproterozoic ocean chemistry: Insights from the Chuar Group, Grand Canyon, USA. *Earth and Planetary Science Letters*, **290**(1-2), 64–73.
- Kaneda T (1991) Iso- and anteiso-fatty acids in bacteria: biosynthesis, function, and taxonomic significance. *Microbiological Reviews*, **55**(2), 288.
- Klepac-Ceraj V, Hayes CA, Gilhooly WP, Lyons TW, Kolter R, Pearson A (2012) Microbial diversity under extreme euxinia: Mahoney Lake, Canada. *Geobiology*,
- Knoll AH, Bambach RK, Payne JL, Pruss S, Fischer WW (2007) Paleophysiology and end-Permian mass extinction. *Earth and Planetary Science Letters*, **256**(3–4), 295–313.
- Logan GA, Hayes JM, Hieshima GB, Summons RE (1995) Terminal Proterozoic reorganization of biogeochemical cycles. *Nature*, **376**(6535), 53–56.
- Lowe DJ, Green JD, Northcote TG, Hall KJ (1997) Holocene Fluctuations of a Meromictic Lake in Southern British Columbia. *Quaternary Research*, **48**(1), 100–113.
- Ludlam SD (1974) Fayetteville Green Lake, New York. 6. The role of turbidity currents in lake sedimentation. *Limnol. Oceanogr*, **19**(4), 656–664.
- Lyons TW, Reinhard CT (2009) An early productive ocean unfit for aerobics. *Proceedings of the National Academy of Sciences*, **106**(43), 18045–18046.
- MacIntyre S, Alldredge AL, Gotschalk CC (1995) Accumulation of marine snow at density discontinuities in the water column. *Oceanography*, **40**(3).



- Madigan MT, Takigiku R, Lee RG, Gest H, Hayes JM (1989) Carbon isotope fractionation by thermophilic phototrophic sulfur bacteria: evidence for autotrophic growth in natural populations. *Applied Environmental Microbiology*, **55**(3), 639–644.
- Meyer KM, Kump LR (2008) Oceanic Euxinia in Earth History: Causes and Consequences. *Annual Review of Earth and Planetary Sciences*, **36**(1), 251–288.
- Meyer KM, Macalady JL, Fulton JM, Kump LR, Schaperdoth I, Freeman KH (2011) Carotenoid biomarkers as an imperfect reflection of the anoxygenic phototrophic community in meromictic Fayetteville Green Lake. *Geobiology*, **9**(4), 321–329.
- Mollenhauer G, Eglinton TI (2007) Diagenetic and sedimentological controls on the composition of organic matter preserved in California Borderland Basin sediments. *Limnology and oceanography*, **52**(2), 558–576.
- Mudge SM, Norris CE (1997) Lipid biomarkers in the Conwy Estuary (North Wales, U.K.): a comparison between fatty alcohols and sterols. *Marine Chemistry*, **57**(1–2), 61–84.
- Naraoka H, Ishiwatari R (2000) Molecular and isotopic abundances of long-chain n-fatty acids in open marine sediments of the western North Pacific. *Chemical Geology*, **165**(1–2), 23–36.
- Nielsen JK, Shen Y, Piasecki S, Stemmerik L (2010) No abrupt change in redox condition caused the end-Permian marine ecosystem collapse in the East Greenland Basin. *Earth and Planetary Science Letters*, **291**(1–4), 32–38.
- Nishihara M, Koga Y (1987) Extraction and Composition of Polar Lipids from the Archaeobacterium, *Methanobacterium thermoautotrophicum*: Effective Extraction of Tetraether Lipids by an Acidified Solvent. *Journal of Biochemistry*, **101**(4), 997–1005.
- Northcote TG, Hall KJ (1983) Limnological contrasts and anomalies in two adjacent saline lakes. *Hydrobiologia*, **105**(1), 179–194.
- Northcote TG, Halsey TG (1969) Seasonal Changes in the Limnology of Some Meromictic Lakes in Southern British Columbia. *Journal of the Fisheries Research Board of Canada*, **26**(7), 1763–1787.
- Otte S, Kuenen JG, Nielsen LP, Paerl HW, Zopfi J, Schulz HN, Teske A, Strotmann B, Gallardo VA, Jørgensen BB (1999) Nitrogen, Carbon, and Sulfur Metabolism in Natural Thioploca Samples. *Applied and Environmental Microbiology*, **65**(7), 3148.
- Overmann J, Beatty JT, Hall KJ, Pfennig N, Northcote TG (1991) Characterization of a dense, purple sulfur bacterial layer in a meromictic salt lake. *Limnology and Oceanography*, **36**(3), 846–859.

- Overmann J, Fischer U, Pfennig N (1992) A new purple sulfur bacterium from saline littoral sediments, *Thiorhodovibrio winogradskyi* gen. nov. and sp. nov. *Archives of Microbiology*, **157**(4), 329–335.
- Overmann, Sandmann G, Hall KJ, Northcote TG (1993) Fossil carotenoids and paleolimnology of meromictic Mahoney Lake, British Columbia, Canada. *Aquatic Sciences*, **55**(1), 31–39.
- Overmann, Thomas Beatty J, Hall KJ (1994) Photosynthetic activity and population dynamics of *Amoebobacter purpureus* in a meromictic saline lake. *FEMS microbiology ecology*, **15**(3), 309–319.
- Overmann J., Beatty JT, Krouse HR, Hall KJ (1996a) The sulfur cycle in the chemocline of a meromictic salt lake. *Limnology and oceanography*, **41**(1), 147–156.
- Overmann J, Beatty JT, Hall KJ (1996b) Purple sulfur bacteria control the growth of aerobic heterotrophic bacterioplankton in a meromictic salt lake. *Applied and Environmental Microbiology*, **62**(9), 3251–3258.
- Overmann J (1997) Mahoney Lake: a case study of the ecological significance of phototrophic sulfur bacteria. In: *Advances in microbial ecology 15* (ed. Jones JG). Springer Science, New York, pp. 251–284.
- Parkes RJ, Taylor J (1983) The relationship between fatty acid distributions and bacterial respiratory types in contemporary marine sediments. *Estuarine, Coastal and Shelf Science*, **16**(2), 173–174.
- Pearson A, McNichol AP, Benitez-Nelson BC, Hayes JM, Eglinton TI (2001) Origins of lipid biomarkers in Santa Monica Basin surface sediment: a case study using compound-specific  $\Delta^{14}\text{C}$  analysis. *Geochimica et Cosmochimica Acta*, **65**(18), 3123–3137.
- Perry GJ, Volkman JK, Johns RB, Bavor Jr HJ (1979) Fatty acids of bacterial origin in contemporary marine sediments. *Geochimica et Cosmochimica Acta*, **43**(11), 1715–1725.
- Planavsky NJ, McGoldrick P, Scott CT, Li C, Reinhard CT, Kelly AE, Chu X, Bekker A, Love GD, Lyons TW (2011) Widespread iron-rich conditions in the mid-Proterozoic ocean. *Nature*, **477**(7365), 448–451.
- Rao Z, Zhu Z, Jia G, Henderson ACG, Xue Q, Wang S (2009) Compound specific  $\delta\text{D}$  values of long chain n-alkanes derived from terrestrial higher plants are indicative of the  $\delta\text{D}$  of meteoric waters: Evidence from surface soils in eastern China. *Organic Geochemistry*, **40**(8), 922–930.
- Schubotz F, Wakeham SG, Lipp JS, Fredricks HF, Hinrichs K-U (2009) Detection of microbial

- biomass by intact polar membrane lipid analysis in the water column and surface sediments of the Black Sea. *Environmental Microbiology*, **11**(10), 2720–2734.
- Scott KM, Henn-Sax M, Harmer TL, Longo DL, Frame CH, Cavanaugh CM (2007) Kinetic isotope effect and bio-chemical characterization of form IA RubisCO from the marine cyanobacterium *Prochlorococcus marinus* MIT9313. *Limnology and Oceanography*, **52**(5), 2199.
- Shulse CN, Allen EE (2011) Diversity and distribution of microbial long-chain fatty acid biosynthetic genes in the marine environment. *Environmental Microbiology*, **13**(3), 684–695.
- Sirevåg R, Buchanan BB, Berry JA, Troughton JH (1977) Mechanisms of CO<sub>2</sub> fixation in bacterial photosynthesis studied by the carbon isotope fractionation technique. *Archives of Microbiology*, **112**(1), 35–38.
- Sorokin II (2002) *The Black Sea: ecology and oceanography*. Backhuys Publishers, Leiden.
- Stein SE (1999) An integrated method for spectrum extraction and compound identification from gas chromatography/mass spectrometry data. *Journal of the American Society for Mass Spectrometry*, **10**(8), 770–781.
- Summons RE, Bird LR, Gillespie AL, Pruss SB, Roberts M, Sessions AL (2013) Lipid biomarkers in ooids from different locations and ages: evidence for a common bacterial flora. *Geobiology*, **11**(5), 420–436.
- Tank M, Thiel V, Imhoff JF (2009) Phylogenetic relationship of phototrophic purple sulfur bacteria according to *pufL* and *pufM* genes. *International Microbiology*, **12**(3), 175–185.
- Taylor J, Parkes RJ (1983) The Cellular Fatty Acids of the Sulphate-reducing Bacteria, *Desulfobacter* sp., *Desulfobulbus* sp. and *Desulfovibrio desulfuricans*. *Journal of General Microbiology*, **129**(11), 3303–3309.
- Velinsky DJ, Fogel ML (1999) Cycling of dissolved and particulate nitrogen and carbon in the Framvaren Fjord, Norway: stable isotopic variations. *Marine Chemistry*, **67**(3–4), 161–180.
- Volkman J (2003) Sterols in microorganisms. *Applied Microbiology and Biotechnology*, **60**(5), 495–506.
- Volkman JK, Burton, HR, Everitt DA, Allen DI (1988) Pigment and lipid compositions of algal and bacterial communities in Ace Lake, Vestfold Hills, Antarctica. *Hydrobiologia*, **165**, 41–57.

- Voorhies AA, Biddanda BA, Kendall ST, Jain S, Marcus DN, Nold SC, Sheldon ND, Dick GJ (2012) Cyanobacterial life at low O<sub>2</sub>: community genomics and function reveal metabolic versatility and extremely low diversity in a Great Lakes sinkhole mat. *Geobiology*, **10**(3), 250–267.
- Wakeham SG (1982) Organic matter from a sediment trap experiment in the equatorial north Atlantic: wax esters, sterol esters, triacylglycerols and alkyldiacylglycerols. *Geochimica et Cosmochimica Acta*, **46**(11), 2239–2257.
- White DC, Davis WM, Nickels JS, King JD, Bobbie RJ (1979) Determination of the sedimentary microbial biomass by extractable lipid phosphate. *Oecologia*, **40**(1), 51–62.

## CHAPTER 3

### Pigments in Mahoney Lake

#### Introduction

Marine oxygen levels have been low during critical periods of Earth's history (Jenkyns, 1980; Brocks et al., 2005; Grice et al., 2005; Lyons et al., 2009). Understanding the impact that anoxia has on chemical and biological processes is therefore important in interpreting the geologic record. The study of meromictic, or permanently stratified, lakes provides one major type of modern analogue for marine paleo-anoxia (Meyer and Kump, 2008). These are lakes where wind or temperature driven mixing of oxygenated water to the lake bottom has been persistently halted, often by density stratification, and the depths of the lake have become anoxic, most commonly euxinic (Findenegg, 1937; Wetzel, 2001; Hall and Northcote, 2012).

The chemoclines of such meromictic lakes often intersect the photic zone, and when this occurs they support populations of anoxygenic photosynthetic bacteria. The primary electron donor for such photosynthesis sometimes is iron (Crowe et al., 2008), but more commonly is sulfide (Jørgensen et al., 1991; Lüthy et al., 2000). Sulfide-driven photosynthesis therefore is a critical link between the carbon and sulfur cycles (Johnston et al., 2009). The most abundant of these organisms are the purple sulfur bacteria (PSB), specifically the families Chromatiaceae and Ectothiorhodospiraceae of the Gammaproteobacteria; and the green sulfur bacteria (GSB) of the

family Chlorobiaceae. Specific members of each of these groups have distinctive pigments, including the isoprenoid carotenoid pigments and bacteria-specific types of chlorophyll. Taxonomically-diverse genera of PSB make okenone, although importantly, okenone is not produced by all PSB and has been suggested to be exclusive to planktonic rather than mat-dwelling or benthic species. (Brocks and Schaeffer, 2008; Takaichi, 2009). However, recently this strict division in the physiological and taxonomic distribution of okenone has been challenged (Meyer et al., 2011). Less common in lakes, but more common in marine systems and in deeper photic zones, the GSB may produce either isorenieratene or chlorobactene (Montesinos et al., 1983; Imhoff, 1995; Ohkouchi et al., 2005). The carbon backbone structures of these carotenoids are known to be geologically well-preserved (Schaeffer et al., 1997) and have been used as proxies for photic-zone euxinia (Summons and Powell, 1986; Brocks et al., 2005; Brocks and Schaeffer, 2008). Additionally, sulfur bacteria do not use the chlorophyll *a* commonly found in oxygenic photoautotrophs, but instead PSB primarily use bacteriochlorophyll *a* while GSB additionally use bacteriochlorophyll *c*, *d*, and *e* (Senge and Smith, 1995; Xiong et al., 2000). Both purple and green sulfur bacteria are found in littoral mat communities and as planktonic species in anoxic water columns (Stal et al., 1985; Caumette et al., 1991; Overmann et al., 1992; Villanueva et al., 1994; Meyer et al., 2011). Understanding the distribution of pigment production in planktonic and benthic species is therefore critical to a proper taphonomic understanding of the geologic record. Euxinic water-columns retard the degradation of organic matter, and carotenoids and porphyrins produced by photosynthetic bacteria and eukaryotes can be highly preserved in lacustrine and marine sediments. This makes them excellent biomarkers to assess climatic and biological changes over long periods of time (Züllig & Rheineck, 1985;

Summons et al., 1986; Sanger, 1988; Repeta et al., 1993; Lowe et al., 1997).

Mahoney Lake is a meromictic lake in British Columbia which has a 10-cm thick layer of PSB in its 7-m deep chemocline. The dominant species in this layer was isolated and identified in culture as *Lamprocystis purpurea* (Overmann et al., 1991). Given new information about its taxonomic affinity, it now is classified as *Chromatiaceae* strain ML1, a member of the genus *Thiohalocapsa* (Tank et al., 2009; Hamilton et al., submitted). The primary carotenoid in these bacteria is believed to be okenone (Overmann et al., 1991; Overmann et al., 1993; Coolen and Overmann, 1998), but rhodovibrin is suggested to be a major pigment in the shoreline microbial PSB species *Thiorhodovibrio winogradskyi*, a genus closely related to *Thiohalocapsa* (Tank et al., 2009). In all cases, these early pigment assignments were assessed only by UV-visible spectrophotometry of lake water and sediment samples and extracts of isolated cultures. The overall pigment composition of both the lake and its bacterial community has not to date been verified by mass spectrometry. . Some of the highest concentrations of bacteriochlorophyll *a* in the world also have been observed in Mahoney Lake (Hall and Northcote, 1990). Here we use multiple extraction procedures and analytical techniques to report pigment abundances from the Mahoney Lake water column, shoreline, and deep basinal sediment. These abundances are useful in investigating pigment production, transport, and degradation processes in euxinic settings. Given the high concentrations of okenone in Mahoney Lake, we also reanalyze previous literature data on okenone, its biosynthetic pathway, and its links to other carotenoid biosynthetic pathways.

## Methods

Mahoney Lake bottom sediment, shoreline mat, and the 7-m purple sulfur bacterial layer were extracted via two different methods. The first iteration used the Bligh-Dyer method (Bligh and Dyer, 1959; Nishihara and Koga, 1987) to give total lipid extracts (TLEs); details are given in Chapter 1. 90% of each TLE obtained from this extraction was separated via silica gel chromatography to yield polarity fractions. Silica gel fractions are labeled with the solvent used to elute each fraction from the column: hexane (A), 5% ethyl acetate in hexane (B), 15% ethyl acetate in hexane (C), 20% ethyl acetate in hexane (D), 25% ethyl acetate in hexane (E), 75% ethyl acetate and 25% methanol (F), and methanol (G). More details of column separation also are given in Chapter 1. All TLEs and polarity fractions were stored at -20°C until analysis.

Because the Bligh-Dyer TLEs originally were generated for fatty acid analysis (Chapter 1) and only later analyzed for pigments, special precautions for pigment analysis were not taken and they may have been subject to excess light, heat, and/or oxidation effects that can cause degradation of labile carotenoids and chloropigments. To determine if this was the case, we re-extracted unweighed aliquots of the 7-m purple sulfur bacterial biomass, the lake bottom sediment, and the shoreline microbial mat by sequentially sonicating in dichloromethane (DCM) and acetone. Each extraction mixture also was modified by the addition of dropwise 12N HCl to remove carbonate and acidify the solutions, thereby also stripping metals from chlorophylls, yielding pheophytins. The DCM and acetone extracts were combined, water removed by Na<sub>2</sub>SO<sub>4</sub>,



and dried under N<sub>2</sub> gas. TLEs were then re-dissolved in acetone and filtered through 0.45-μm PTFE syringe filters before analysis. All steps were performed in darkened hoods and extracts were stored at 4°C and analyzed within one week of extraction.

Extracts in acetone were analyzed on an Agilent model 1290 ultra-high pressure liquid chromatograph with model 6410 triple quadrupole mass spectrometer (UHPLC-QQQ) running Agilent Masshunter version C.01.05. Unless otherwise stated, all retention times, UV spectra, and other peak information refers to observations from these chromatograms.

For high-resolution chromatography, we followed the “A” method of Airs et al. (2001). Because our samples were not previously methylated with diazomethane, we used 0.5 M ammonium acetate as the ion-pairing reagent A as recommended in that method. Solvents B, C, and D were methanol, acetonitrile, and ethyl acetate, respectively, and the gradient proceeded from 5/80/15/0% to 0/20/15/65% A/B/C/D over 95 minutes (initial isocratic hold, 5 min). The mobile phase then transitioned to 0/1/1/98% A/B/C/D over 5 minutes, was held isocratically (5 min), and then transitioned back to the starting composition of 5/80/15/0% A/B/C/D. The flow rate was 0.5 ml min<sup>-1</sup> and three Kinetex C<sub>18</sub> phase (150 x 4.6 mm, 2.6 μm particles) columns were used in series. Each run was analyzed by ultraviolet-visible diode array detector (DAD) and by tandem mass spectrometry with atmospheric pressure chemical ionization (APCI) in positive ion mode. Ionization parameters were chosen to reduce in-source fragmentation of okenone: vaporizer temperature 300°C, drying gas flow 4 L/min at 300°C, nebulizer pressure 40 psi, fragmentation 135 V, and collision chamber 20 V. The mass spectrometer was operated in two modes: by

monitoring specific ion fragmentation transitions for compounds that may be present in the sample (Table 3.1), and by scanning all ions from 100 to 1000 amu. Each sample was analyzed twice, once with each mode.

In addition to identifying compounds by relative retention time (according to Airs et al., 2001; Hodgson et al., 2004; Romero-Viana et al., 2009), they were identified by their UV-visible spectra and mass spectral fragmentation patterns, including molecular ion masses ( $M^+$ ) and  $MS^2$  transitions (Table 3.1). Pure standards were analyzed to determine the most abundant daughter products for  $MS^2$  transitions of beta-carotene and retinoic acid. The sediment TLE was analyzed to determine the transitions for okenone, bacteriopheophytin *a*, and pyrobacteriopheophytin *a*. A TLE extracted from *Allochromatium vinosum* was analyzed to determine the transition for rhodopin, and a TLE extracted from *Chlorobium limicola* was analyzed to determine the transition for chlorobactene. As echinenone contains neither a hydroxy nor a methoxy group, its transition was chosen from literature observations of its major fragmentation ions (Airs et al., 2001). The best transitions for all other compounds were determined by analogy to known transitions of compounds with the same functional groups.

Table 3.1: Transitions monitored during high-resolution QQQ runs of acetone extracts. Stars (\*) indicate transitions seen in analysis of previously identified compounds; Daggers (†) indicate transition drawn from literature; all others inferred.

Compound(s)	Transition	Moiety Lost
Bacteriopheophytin <i>a</i> *	889.6→611.4	Phytol
Pyrobacteriopheophytin <i>a</i> *	831.5→552.3	Phytol
Phaeophytin <i>a</i>	871.5→593.3	Phytol
Lutein / Zeaxanthin	569.4→551.4	OH
Canthaxanthin / Alloxanthin	565.4→547.4	OH
Echinenone <sup>†</sup>	551.4→459.4	C <sub>7</sub> H <sub>8</sub>
Fucoxanthin	659.4→641.4	OH
Okenone*	579.5→547.5	MeO
Thiothece-484	623.4→591.4	MeO
Spheroidenone / <i>R.g.</i> keto 1 / Thiothece-474	583.4→551.4	MeO
Thiothece-460	463.4→431.3	MeO
Thiothece-OH-484	601.4→583.4	OH
<i>R.g.</i> keto 2	613.4→581.4	MeO
<i>R.g.</i> keto 3	629.4→596.4	MeO
Spheroidene	569.4→537.4	MeO
Rhodopin* / Demethylspheroidene	555.5→537.4	OH
Anhydrorhodovibrin	567.5→535.4	MeO
Rhodovibrin	585.5→567.5	OH
Spirilloxanthin	597.5→565.6	MeO
Chlorobactene*	533.4→133.0	Benzyl Ion
Isorenieratene	529.4→133.0	Benzyl Ion
β-carotene* / Lycopene	537.5→445.4	C <sub>7</sub> H <sub>8</sub>

To quantify compound abundances, some samples were analyzed using a substantially shorter version of the Airs et al. (2001) method. For these runs, a retinoic acid internal standard was added to the samples, and its response was compared to an external calibration curve having concentrations of retinoic acid of 1 ng, 10 ng, 100 ng, 1 µg, and 10 µg (run in triplicate). Peak areas (QQQ total ion intensity) of compounds from the sample runs were then converted to concentrations by this external curve. The response was non-linear above concentrations of 500 ng (peak areas of 36,000 counts x min); concentrated samples were diluted so that their peak areas would fall below this threshold. The detection limit estimated from the retinoic acid response curve was 30 pg.

The short method (20 minutes versus 115 minutes in Airs et al., 2001) is adequate for quantification, as compounds need not be fully separated given the specificity of QQQ-APCI-MS<sup>2</sup> detection (also known as MRM, or multiple reaction monitoring). The short method begins with 3/50/15/32% A/B/C/D and ramps to 0/10/15/75% A/B/C/D over 10 minutes after a 2 minute isocratic hold. The column (Zorbax Eclipse XDB-C<sub>18</sub>, 4.6 x 150 mm, 5 µm particles) is then flushed with 0/2/0/98% A/B/C/D for 8 minutes to elute any strongly nonpolar compounds. Ionization parameters were as above, but to promote ionization the vaporizer temperature was raised to 450°C and collision energies for carotenoid species—excluding the unfunctionalized carotenoids beta-carotene, isorenieratene, and chlorobactene—were lowered to 10 V to prevent excessive fragmentation. Only a subset of the transitions used for the long method was monitored with this method (Table 3.2). This method was used to quantify pigment compound abundances in the total lipid extracts from the Bligh-Dyer extractions. It was not applied directly

to the acetone-extracted aliquots. Instead, the concentration of bacteriopheophytin *a* from each sample was assumed to be the same in both the Bligh-Dyer and the acetone extracts, and all remaining compounds in the acetone extracts were scaled accordingly to obtain their *in situ* concentrations.

*Table 3.2: Carotenoids and their associated fragmentations patterns scanned during quantitative QQQ runs. Stars (\*) indicate transitions seen in analysis of previously identified compounds; all others inferred. Dagger (†) indicates transition was misentered as 536.4→444.4.*

Compound	Transition	Moiety Lost
Bacteriopheophytin <i>a</i> *	889.6 → 611.4	Phytol
Pyrobacteriopheophytin <i>a</i> *	831.5 → 552.5	Phytol
Lutein / Zeaxanthin	569.4 → 547.5	OH
Okenone*	579.4 → 547.5	MeO
Rhodopin*	555.5 → 537.4	OH
Anhydrorhodovibrin	567.5 → 535.4	MeO
Rhodovibrin	585.5 → 567.5	OH
Spirilloxanthin	597.5 → 565.6	MeO
Chlorobactene*	533.4 → 133.0	Benzyl ion
Isorenieratene	529.4 → 133.0	Benzyl ion
β-carotene*	537.4 → 445.4 †	C <sub>7</sub> H <sub>8</sub>
Retinoic acid*	301.4 → 283.3	OH

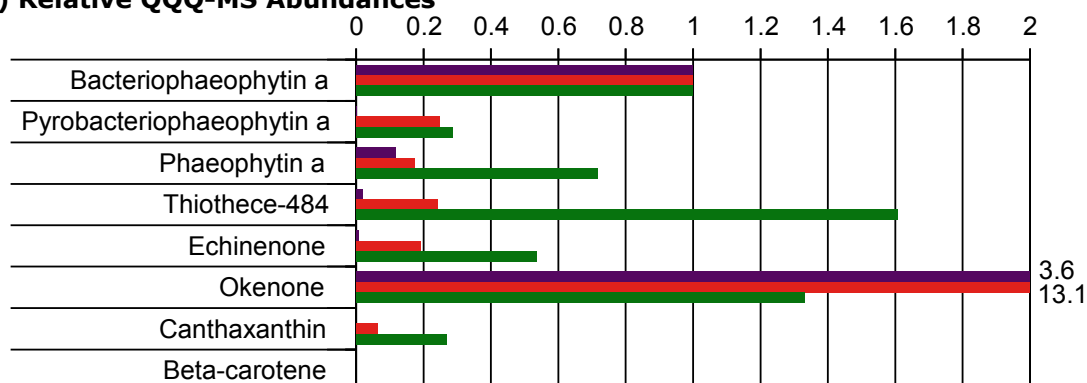
Relative concentrations of pigments were also determined in several alternative ways for the acetone extracts. Mass spectral peak areas from extracted ions representing full-spectrum scans from 100 to 1000 amu were measured. Additionally, UV-vis spectral peak areas were converted to concentrations using Beer's Law, as modified for use with HPLC (Torsi et al., 1990) according to:  $N = (A * F) / (\epsilon * l)$  where N is the number of moles, A is the peak area,  $\epsilon$  is the molar

absorptivity and  $l$  is the spectrometer cell length. Because the injections represent an unknown fractional quantity of the initial extracts, each of these approaches then reports the data normalized to the most consistently abundant peak, bacteriopheophytin. All peak areas were determined by integration in Agilent Masshunter. UV-visible chromatograms then were graphically transformed to red-green-blue colors suitable for display using analytic approximations of the CIE 1931 color space (Wyman et al., 2013) and shown as simulated chromatographic columns. To reveal low-intensity bands, contrast of these images was increased using the GNU Image Manipulation Program ([www.gimp.org](http://www.gimp.org)).

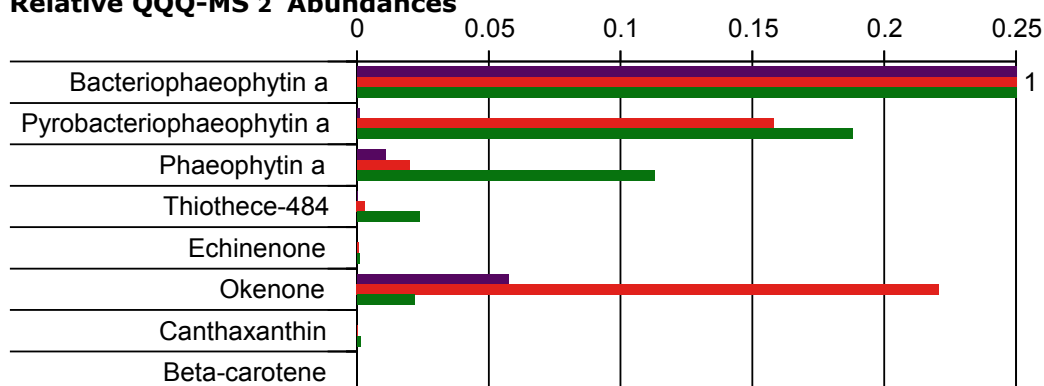
Finally, several presumed biosynthetic intermediates and/or diagenetic breakdown products of the okenone and *R.g.* keto carotenoid biosynthetic series (see below) were detected in several samples. To gain confidence in their tentative structural assignments, silica gel fractions from the Bligh-Dyer extractions were run to check compound polarity and relative retention times in comparison to other structural analogues and prior literature (Andrewes and Liaaen-Jensen, 1972; Overmann et al., 1993; Britton et al., 2004). Fractions C, D, E, and F were injected on an Agilent 1100 series HPLC with single quadrupole MSD following the method of Barua and Olson (1998). The column was a Zorbax Eclipse XDB-C18 4.6 x 150 mm (flow rate 0.5 ml min<sup>-1</sup>). After a one minute isocratic hold with 10 mM ammonium acetate in 3:1 methanol:water (solvent A), the HPLC program followed a gradient to 4:1 methanol:dichloromethane (solvent B) over 15 min, and then was isocratic for 30 min. Column cleaning was achieved by back-flushing with isopropanol (10 min) followed by a re-equilibration in solvent A (10 min). UV-visible spectra (250 nm to 800 nm; 2 nm steps, 4 nm slit width) and atmospheric pressure chemical

ionization (APCI) mass spectra from 100 amu to 1000 amu were collected. Again optimizing for the best response of two standards, retinoic acid and beta-carotene, positive APCI was used with fragmentor voltage 175 V, drying gas flow 6 L/min @ 350°C, vaporizer temp 375°C, nebulizer pressure 60 psi, capillary voltage 3000 V, and corona current 5  $\mu$ A. Runs were bracketed by injections of 500 ng retinoic acid; response from this external standard was used to semi-quantitatively estimate compound abundances in addition to UV-visible absorbances, as above.

**(A) Relative QQQ-MS Abundances**



**(B) Relative QQQ-MS 2 Abundances**



**(C) Relative UV Abundances**

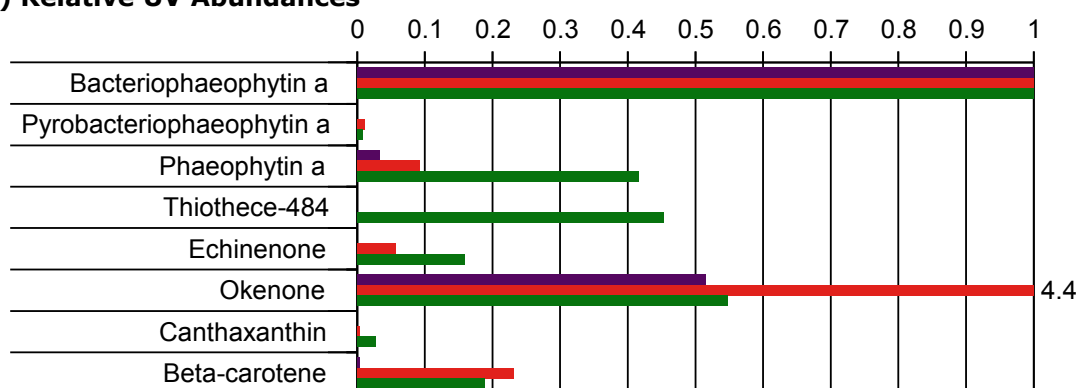
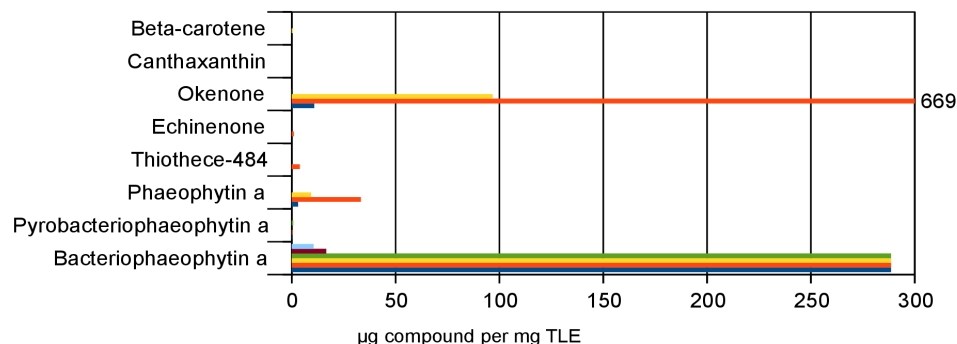


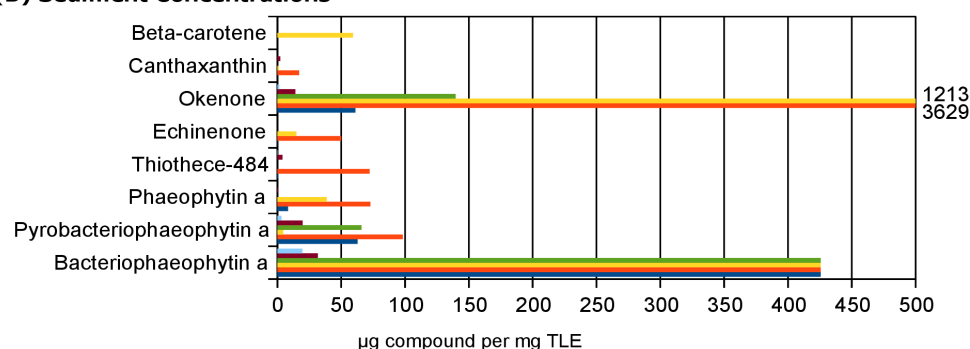
Figure 3.1. Relative abundances of carotenoids and chlorophyll derivatives from acetone extracts: (a) extracted MS ion data, (b) MRM transition data, and (c) UV-visible data. All values are reported as ratios relative to bacteriopheophytin a.



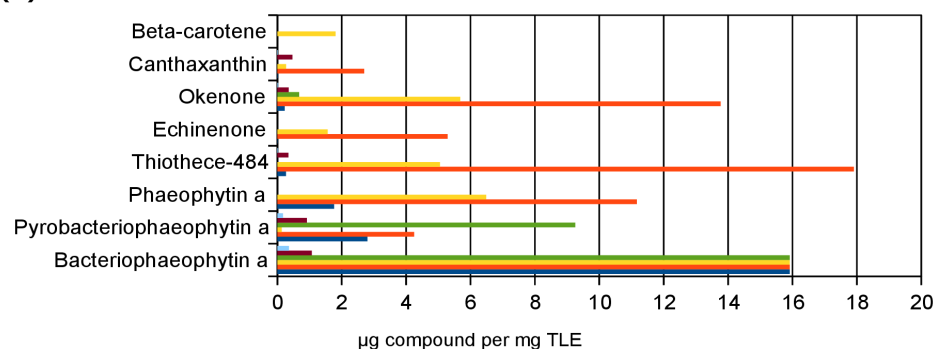
**(A) 7m Concentrations**



**(B) Sediment Concentrations**



**(C) Shoreline Concentrations**



■ Acetone QQQ-MSMS    ■ Acetone QQQ-MS    ■ Acetone UV  
■ Bligh-Dyer QQQ-MSMS    ■ Bligh-Dyer Fxn MS    ■ Bligh-Dyer Fxn UV

Figure 3.2: Semi-quantitative measurements of pigment abundance in Mahoney Lake samples. Quantifications are relative to retinoic acid (assumed relative response factor = 1), and are scaled between different extraction methods assuming each method should yield the same result for bacteriopheophytin a. Such an assumption, however, yields the impossible result of >100% okenone for some samples (> 1 mg/mg TLE), implying that this assumption is inaccurate and that the approach must be only semi-quantitative.

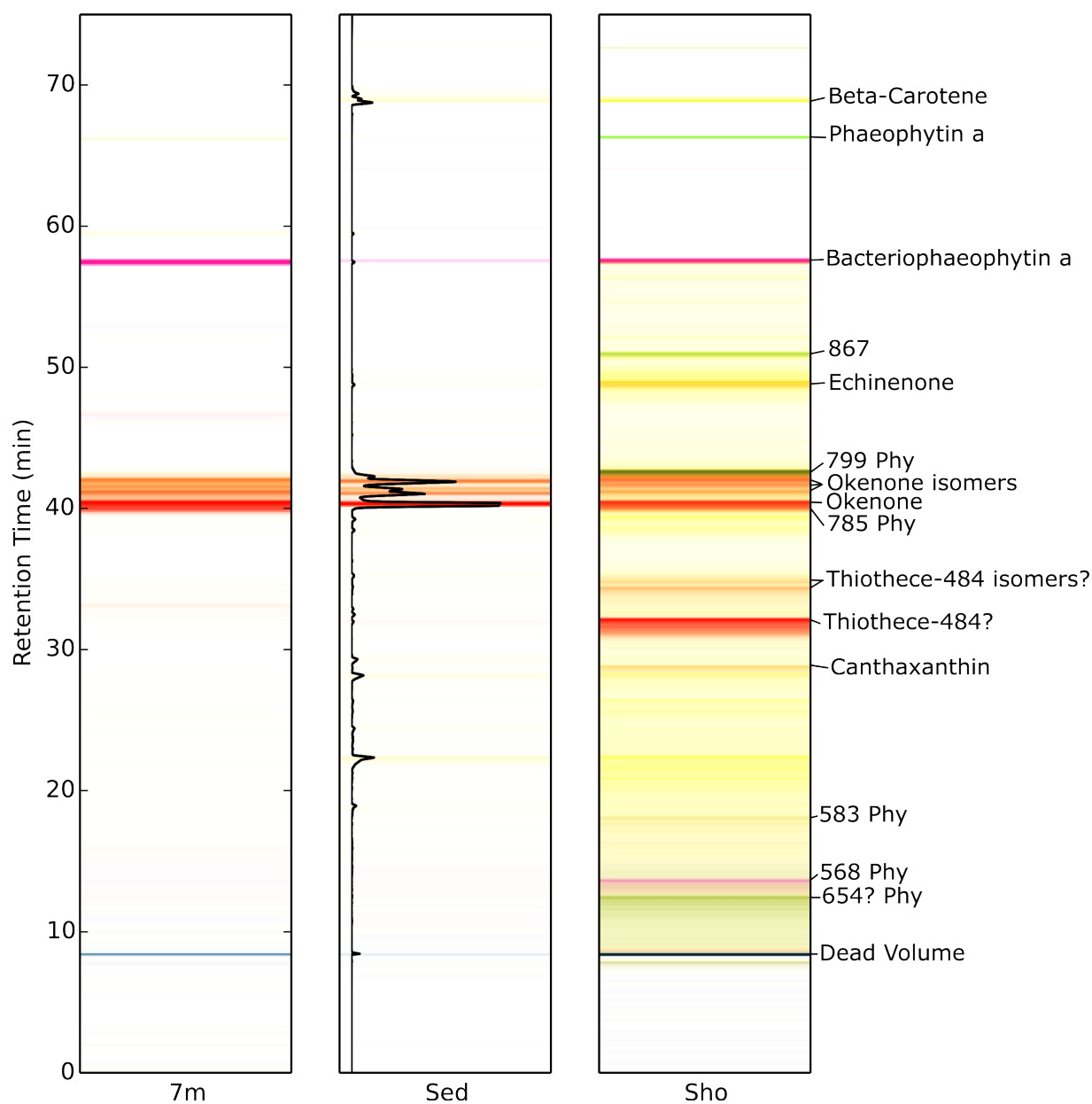


Figure 3.3: UV-visible spectrophotometry of the long runs of acetone extracts with the *Airs et al.* (2001) method presented as simulated columns. Compounds labelled with their  $[M+H]^+$  and “Phy” are believed to be porphyrin derivatives on the basis of their UV-visible spectra. Overlaid on center column is the UV-visible trace at 440 nm to show correlations between colors and visible wavelength peaks of carotenoids.

## Results

Bacteriopheophytin *a* was common across all samples (sediment, shoreline, and 7-m chemocline) and was present in both the acetone and Bligh-Dyer extracts, so we report the concentrations of all other compounds relative to it (Figure 3.1). We also transform these relative values to quantitative estimates by assuming bacteriopheophytin *a* has the same absolute concentration ( $\mu\text{g}/\text{mg}$  TLE) in both the Bligh-Dyer and acetone extracts (Figure 3.2). This compound is one of the primary peaks in both the 7-m and shoreline samples, while it appears to have slightly lower relative abundance in the sediment (Figure 3.3). In UHPLC-QQQ analysis of acetone extracts it eluted at 57.5 minutes, had a  $[\text{M}+\text{H}]^+$  ion of 889.6, and a UV-visible spectrum with maxima at 358, 528, and 748 nm. The same peak was observed in fraction E of the Bligh-Dyer extracts at 29.3 minutes using the chromatographic method of Barua and Olson (1998). An additional low-intensity peak also was observed to have the same mass and UV-visible spectrum, but a retention time of 60 minutes (by method of Airs et al., 2001) or 31 min (by method of Barua and Olson, 1998). This latter peak likely corresponds to the bacteriopheophytin *a* epimer described by Airs et al. (2001).

In addition to bacteriopheophytin, other porphyrin derivatives also were present. A peak at 66.2 minutes (by method of Airs et al., 2001) or in fraction E at 39.1 minutes (by method of Barua and Olson, 1998) with  $[\text{M}+\text{H}]^+$  ion of 871.5 and UV-visible maxima of 408 and 666 nm, followed by a peak at 68.1 minutes with similar characteristics, were identified as phaeophytin *a* and phaeophytin *a* epimer (Figure 3.3). They are relatively more abundant in the shoreline than in

any other sample (Figure 3.1). Another peak at 64.0 minutes (in fraction E at 34.0 minutes; Barua and Olson, 1998) with an  $[M+H]^+$  ion of 831.5 and UV-visible maxima of 358, 530, and 750 nm was identified as pyrobacteriopheophytin *a*. It also was relatively more abundant in the shoreline than in the water column, but its abundance in the lake bottom sediment was similar to the shoreline (Figure 3.1).

Okenone was the most abundant carotenoid present in both the purple layer and the sediment, and one of the most abundant carotenoids in the shoreline. It was detected with an  $[M+H]^+$  ion of 579.4 and UV-visible spectral maxima of 488 and 514 nm in all samples at a retention time of 40.3 minutes (Figure 3.3). It also was observed in fraction D of the shoreline and sediment Bligh-Dyer extracts at 26.4 minutes with the Barua and Olson method. Although we do not have an authentic standard for comparison, these characteristics match the literature properties of okenone (e.g., Britton et al., 2004; Romero-Viana et al., 2009). Immediately following the okenone peak at retention times of 41.9, 42.3, 42.9, and 43.2 were four other discernible peaks also with  $[M+H]^+$  ions of 579.4 and slightly shifted UV-visible spectral maxima of 480 and 508 nm. A shorter maximum absorption wavelength is characteristic of *cis* carotenoids (Liaaen-Jensen and Lutnæs, 2008). As such, these later peaks likely are *cis* isomers of okenone. In the acetone extracts, peak area ratios of *trans*-okenone to the sum of the *cis* isomers are between 0.7 and 1.7 regardless of sample or quantification technique. This ratio is higher for Bligh-Dyer extracts (1.8 and 2.3 for the relative MS peak areas in Bligh-Dyer silica gel fractions), but possible sample handling-related degradation effects and the failure to detect okenone in the 7-m sample extracted by this method (Figure 3.2a) cautions against interpreting these ratios to have

primary significance in any sample. All ratios are lower than previously reported for the ML1 culture (5.7; Overmann et al., 1993). Although some anoxygenic photoautotrophs produce *cis* carotenoids (Bialek-Bylka et al., 1998), isomerization is one of the most common laboratory artifacts in carotenoid analysis (Liaaen-Jensen, 2004) and all of our low values likely result from the laboratory and are not authentic.

In the shoreline, the most abundant (by both mass spectral techniques, but not by UV-visible spectrum; Figure 3.1) carotenoid eluted at 32.0 minutes with an  $[M+H]^+$  ion of 623.3 and a fragment at 591.2; UV-visible maxima of 490 and 518 nm; and a small secondary isomer at 34.8 minutes with identical mass and similar spectrum. This peak was also observed using the Barua and Olson method, eluting at 23.1 minutes in fraction D from the shoreline and fraction E from the sediment. Its presence in different polarity fractions is presumably due to minor inconsistencies in  $\text{SiO}_2$ -gel chromatography between samples, but the tendency toward fraction E suggests it is more polar than okenone. This interpretation also is supported by the shorter relative retention time in reverse-phase HPLC (Figure 3.3). A mass loss of 32-amu is similar to the fragmentation for okenone and likely corresponds to loss of a methoxy group. Based on a molecular mass of 622, the presence of at least one methoxy group, a UV-visible spectrum similar to okenone, and a slightly greater polarity than okenone, we identify this peak as the carotenoid Thiothece-484 (Andrewes and Liaaen-Jensen, 1972). Lesser amounts of Thiothece-484 were observed in the sediment and very little was observed in the 7-m chemocline sample (Figures 3.1, 3.2).

Two other carotenoids were found in notable abundances. A peak at 28.5 minutes with an  $[M+H]^+$  ion of 565.4 and a UV-visible maximum of 474 was abundant in the shoreline. This peak also was seen in fraction E of the shoreline at 21.6 minutes by the method of Barua and Olson. Both alloxanthin and canthaxanthin have mass 564, but canthaxanthin has a UV-visible maximum of 474 nm (in ethanol) while alloxanthin has its maximum at 450 nm (in ethanol) and greater fine structure with additional maxima at 427 and 478 nm (Britton et al., 2004). On the basis of these spectral differences, we assign this peak to canthaxanthin. Another peak at 49 minutes with an  $[M+H]^+$  ion of 551.4 and a UV-visible maximum of 482 nm was observed in the acetone extracts (Figure 3.1), but not in the Bligh-Dyer extracts (Figure 3.2); it is identified tentatively as echinenone, a structural relative of canthaxanthin having only a single keto-group. Both were relatively more abundant in the shoreline than in the basinal sediment, and they were detected at only trace levels (echinenone) or not at all (canthaxanthin) in the 7-m chemocline sample (Figure 3.1).

Surprisingly, several compounds which might be expected to be found were not observed using any of the approaches. Rhodopin is reported to be the primary carotenoid in *Thiorhodovibrio winogradskyi*, a species cultured from the Mahoney Lake shoreline (Overmann et al., 1992). However, we found no definitive peaks of any of the spirilloxanthin series of carotenoids—including rhodopin—in any of the samples. Several small peaks having transitions 585.5→567.5 (hypothetical transition of rhodovibrin) and 555.5→537.4 (transition of rhodopin) were detected

in the MRM trace of the shoreline sample at ca. 30 minutes (by method of Airs et al., 2001), but examination of the full MS<sup>2</sup> trace shows these to be in-source degradation products of phytins. Additionally, despite the reported presence of Chrysophyte algae in the epilimnion (Northcote and Hall, 1983), fucoxanthin was not found in any of the samples (nor was it detected in earlier carotenoid work on sediments; Overmann et al., 1993).

Results from the Bligh-Dyer extracts differed significantly from the acetone extracts. These differences may be attributed to loss on SiO<sub>2</sub>-gel, degradation during processing, and/or different inherent extraction yields. The most abundant carotenoid observed in the 7-m layer Bligh-Dyer extract eluted in silica gel fraction E, had a mass spectrum with an [M+H]<sup>+</sup> ion at *m/z* 463, an abundant mass fragment at *m/z* 431 and UV maxima at 480 and 506 nm. Three peaks of this compound were resolved, presumably indicating three isomers. Due to its parent mass of 462, the presence of a M-32 ion, and the observed spectral maxima at 474 and 500 nm, we tentatively identify this compound as Thiothece-460 (Britton et al 2004). The other major carotenoid in the 7-m sample was found in fraction D. It had an [M+H]<sup>+</sup> ion at *m/z* 609 (primary fragments at 577 and 545) with UV-Visible spectral maxima at 460 and 488 nm. Several peaks with an [M+H]<sup>+</sup> ion of 595 also were very abundant in fractions D and E from the sediment, and were present in those same fractions in the shoreline and 7-m samples. Some of these peaks had MS fragmentation patterns showing abundant M-16 and M-32 ions, and their spectral maxima were consistently at 446 and 468 nm. However, upon looking for all of these compounds in the acetone extracts, only trace amounts of Thiothece-460 were observed and none of the other unusual carotenoids were found. This suggests the Bligh-Dyer samples may have experienced

significant degradation during processing and the above compounds may be artifacts. Such degradation may have affected both the carotenoids and the chloropigments.

We attempted to determine the quantitative abundances of both classes of pigments ( $\mu\text{g}$  compound per mg TLE) in all samples (Figure 3.2). However, this approach relies specifically on the quantification of the Bligh-Dyer extracts, because only these TLEs were weighed. For quantification of the acetone extracts, we assumed a constant concentration of bacteriopheophytin *a* across both extract types and normalized the acetone values to the Bligh-Dyer values. This implies that due to prior degradation of the Bligh-Dyer extracts, the calculated concentrations for all lipids in the acetone extracts are too high. In particular, at  $> 1\text{ mg/mg}$ , the calculated concentrations of okenone in the sediment acetone extract (Figure 3.2b) are physically impossible. However, the concentrations measured for the sediment Bligh-Dyer extract do generally agree with previous observations. Hall and Northcote (1990) measured 1.5 to 7  $\text{mg L}^{-1}$  bacteriochlorophyll *a* in the purple layer. From 500 ml of centrifuged 7 m water yielding 3.48 mg TLE, we measured 1.0 mg of bacteriopheophytin *a* (the acidified, metal-free form of bacteriochlorophyll *a*). Overmann et al. (1993) also measured between 0.1 and 2 mg okenone per gram dry sediment. We extracted 5.20 g wet sediment to yield 3.76 mg TLE, and a previous sample of the same material had a dry mass of 328 mg for a 5.00 g wet weight. Scaled accordingly, our measured 0.5 mg of okenone is equivalent to 1.5 mg okenone per gram sediment.



There are clear indications that instrumental response factors differ for the various chloropigments and carotenoids analyzed in this study. For example, the ratio of okenone to bacteriopheophytin *a* determined by HPLC-QQQ analysis (Figure 3.1a) is greater than what is measured from their respective UV-visible spectra (Figure 3.1c), and this in turn is greater than what is measured by HPLC-QQQ-MS<sup>2</sup> (Figure 3.1b). With respect to the UV method, the discrepancies likely are due to a combination of unpredictable degradation of fluorophores and inherent errors in defining molar absorptivities ( $\epsilon$ ; Torsi et al, 1990). For the mass spectral data, differences in results between methods likely are due to variations in behavior of ions in the collision chamber, reflecting the heterogeneity and concentration differences between the samples. Due to these effects, we aim to be conservative and regard all results as only semi-quantitative; the most robust comparisons should be ratios of the same or very similar compounds both between and within samples, e.g., okenone vs. echinenone or bacteriopheophytin *a* vs. pheophytin *a*.

Both types of extracts did give similar results for the sediment sample. But for both the shoreline mat and the 7-m layer, there were significant differences, the primary one being a failure to detect large quantities of okenone in the Bligh-Dyer samples (Figure 3.2). Also, lutein and zeaxanthin abundances (Appendix 1, Table S4) were much lower than reported previously by Overmann et al. (1993). The Bligh-Dyer extracts showed a poorly-resolved but distinct peak having the expected 569 $\rightarrow$ 547.5 transition; it was detected and quantified under the assumption it is an unresolved mixture of lutein and zeaxanthin. There were no identifiable peaks with an [M+H]<sup>+</sup> ion of *m/z* 569.4 in the acetone extracts.

Beta-carotene also appeared to be present in all samples, but it could not be quantified by mass spectrometry with the approaches used here (Figure 3.3). Peaks tentatively assigned to beta-carotene could be observed in the sediment sample by quantitative UHPLC-QQQ mass scanning (Figure 3.1a), but the incorrect mass— $M^+$  (536.4) instead of the  $[M+H]^+$  ion (537.4); Table 3.2—was inadvertently monitored, so its abundance could not be quantified. When UHPLC-QQQ- $MS^2$  runs were performed, three large UV-Visible peaks were observed starting at 68.8 minutes, but the corresponding mass spectral traces were very subdued and ionization is believed to have been very poor. Due to its lack of heteroatoms, beta-carotene would be expected both to elute very late in reverse-phase HPLC and to have poor ionization potential. These peaks had UV-Visible maxima at 454 and 480 nm, 448 and 474 nm, and 446 and 470 nm. Three peaks with similar retention times and UV-Visible spectra are identified as “carotene” by Airs et al. (2001), consistent with our assignment; we report abundances of beta-carotene only from spectral absorbance (Figure 3.2).

## Discussion

### *Abundance and distribution of pigments in Mahoney Lake samples*

As expected, we find high concentrations of okenone in all samples of Mahoney Lake (Overmann et al., 1991; 1993). Our confirmation of okenone by relative retention time, spectral absorbance, and mass fragmentation patterns suggests that very early reports of rhodovibrin being a major pigment in this environment (Northcote and Hall, 1969) likely were erroneous.

Although there may be temporal patterns in production, Mahoney Lake has been for stratified for more than 9000 years (Lowe et al., 1997), and okenone is the major pigment identified in sediments throughout this interval (Coolen and Overmann, 1998). Secondary stratification develops and disappears in the Mahoney Lake epilimnion on a seasonal cycle (Northcote and Hall, 1990), and the anoxygenic phototrophic communities in other meromictic lakes can change both on seasonal (Bosshard et al., 2000) and decadal (Tonolla et al., 2005; Storelli et al., 2013) timescales, but Mahoney Lake appears to be unusually stable (Overmann et al., 1996a; Overmann, 1997). Okenone is greatly enriched in basinal sediment, both in absolute and relative terms. This appears to reflect the high preservation potential of aromatic carotenoids relative to other pigments (Koopmans et al., 1996; Brocks and Schaeffer, 2008), because both its absolute and relative abundance increases dramatically between the zone(s) of production (7-m and shoreline) and the sediments. Porphyrin breakdown products also change their relative concentrations across samples. Phaeophytin *a*—the metal-free breakdown product of chlorophyll *a*—is most abundant in the shoreline and represents a greater input of either eukaryotic or cyanobacterial primary production to the shoreline. Greater concentrations of pyrobacteriophageophytin *a* in both the shoreline mat and lake bottom sediment, relative to the 7-m sample, indicate the inception of early diagenesis. The presence of significant phaeophytin *a* in the sediments may be an additional tracer for the input of shoreline-derived material (Chapter 1, this thesis). Due to these degradation and transport effects, it is difficult to make quantitative interpretations from the pigment abundances measured here.

Similarly, previous work has exposed a lack of correlation between the abundance of okenone,

bacteriochlorophyll *a*, and DNA sequences of PSB in Mahoney Lake sediments (Coolen and Overmann, 1998). In this work we made semi-quantitative measurements of okenone relative to the abundance of bacteriochlorophyll *a* breakdown products (Figures 3.1, 3.2). These measurements were not quantitative, because we lacked authentic standards to calibrate the mass spectral response factors of bacteriochlorophyll derivatives and okenone, specifically with respect to ionization and collision-chamber fragmentation. However, changes in relative ratios of these compounds are useful both for understanding how the occurrence/disappearance of pigments relates to the physiology of purple sulfur bacteria (Smith et al., 2014); as well as to gain insight into relative degradation rates (e.g., Steenbergen et al., 1994), as multiple degradative pathways exist for both porphyrins (Leavitt, 1993) and carotenoids (Hebting, 2007). Better understanding of both of the biotic and abiotic controls on relative pigment abundances may help explain the lack of correlation with DNA abundances (Coolen & Overmann, 1998) and further constrain organic export to the sediments (Chapter 1, this thesis).

Outside of the dominance of okenone, there also was a wide diversity of minor pigments in the Mahoney Lake samples, including previously detected compounds and novel compounds. In addition, various pigments previously mentioned in the literature could not be found in our samples: Lutein and zeaxanthin were not detected (Appendix 1, Table S4) despite being reported in high relative abundance in Overmann et al. (1993). This may be due to extraction biases, sample degradation, or prior misidentification. Overmann et al. (1993) used an aggressive extraction in 60°C methanol and KOH which may have resulted in improved recovery of these compounds. Conversely these authors did not detect echinenone, a compound which also was not

present in our Bligh-Dyer extract, but which was present in our (gentlest) acetone extract. This suggests that echinenone is particularly fragile and may not survive harsh conditions. However, unlike situations where differences in analytical methodology have caused different environmental interpretations (Coolen and Overmann, 2007; Sinninghe Damsté and Hopmans, 2008), we do not believe that differing identification of these particular compounds substantially changes either the present or previous conclusions about Mahoney Lake.

#### *Identification of unusual carotenoids in Mahoney Lake samples*

Overmann et al. (1993) identified a compound they called “demethylated okenone” and verified its coelution with a pigment present in an extract of the PSB species *Thiocystis gelatinosa*. Although this “demethylated okenone” originally was thought to be okenone possessing a terminal alcohol instead of methoxy moiety (mass=564; Pfennig et al., 1968), later analysis showed this compound to be the structure known as Thiothece-OH-484 (mass=600; Andrewes and Liaaen-Jensen, 1972). Thiothece-OH-484 is a carotenoid that has the same 4'-ketolated  $\psi$  (acyclic) end as okenone, but which has a second, hydroxylated  $\psi$  end instead of the  $\chi$ -ring (Figure 3.4). This compound also is found in the *R.g.* keto pathway of purple non-sulfur bacteria (Schmidt and Liaaen-Jensen, 1973). It has an absorption spectrum similar to that of okenone and the carbomethoxylated okenone derivative Thiothece-484 (Andrewes and Liaaen-Jensen 1972), hence its prior assignment as “demethylated okenone”. Given their functional and absorptive similarities, it is possible that without mass spectral data or an authentic standard for co-elution, Thiothece-484 and Thiothece-OH-484 could be confused. Here we positively identified okenone

and Thiothece-484, but not Thiothece-OH-484 in our samples.

Thiothece-484 has not been reported previously in any environmental sample, although several peaks have been detected with very similar properties. Villanueva et al. (1994) reported an unidentified carotenoid with mass of 622 and spectral maxima of 452 and 474 nm in an evaporite mat also containing spirilloxanthin and rhodopin, but without reported okenone. When reanalyzed with a different method, the same mat had a peak eluting between canthaxanthin and okenone with a  $[M+H]^+$  of 623 and spectral maxima of 381, 460, 483, and 514 in addition to containing detectable okenone (Airs et al., 2001). If these peaks were Thiothece-484, it is possible that it commonly co-occurs with okenone, as seen in the Mahoney Lake samples. Most critically, Thiothece-484 has been found in several cultures of Chromaticeae including *Thiocapsa marina* (Caumette et al., 1985; Caumette et al., 2004) and *Thiohalocapsa halophila* (Caumette et al., 1991). *Thiohalocapsa halophila* is the nearest sister group to Chromaticeae strain ML1 (Tank et al., 2009; Hamilton et al., submitted). It was isolated from a benthic mat on the coast of France (Caumette et al., 1991). Its major carotenoid is okenone, although it does produce spirilloxanthin-series carotenoids in addition to okenone-series carotenoids and Thiothece-484.

It remains unknown whether Thiothece-484 is an intentional biosynthetic end-product, a biosynthetic intermediate, or a diagenetic marker. Both its function and more generally, other possible functional distinctions between okenone, Thiothece-484, or spirilloxanthin-series

carotenoids in PSB are poorly understood. Okenone has light absorption properties that allow Chromaticeae to live deep in the photic zone, near the diffusional supply of sulfide (Imhoff, 2006). Changes in accessory carotenoids may then reflect adaptations to different light intensities. Caumette et al. (1985) observed that as light flux increased from 400 to 2000 lux, the percent Thiothece-484 of total carotenoids in *Thiocapsa marina* cultures increased from 11 to 15% while okenone decreased from 78 to 71% (Caumette et al., 1985); also suggesting this compound is a deliberate synthetic product. The position of Thiothece-484 in the okenone pathway is poorly understood, but could resolve questions of its taxonomic and environmental provenance (Figure 3.4). Initially, it was proposed as a biological oxidation product of okenone (Andrewes and Liaaen-Jensen 1972), while later authors interpreted high concentrations in culture as indicative of a biosynthetic intermediate formed during ring rearrangement (Caumette et al., 1985).

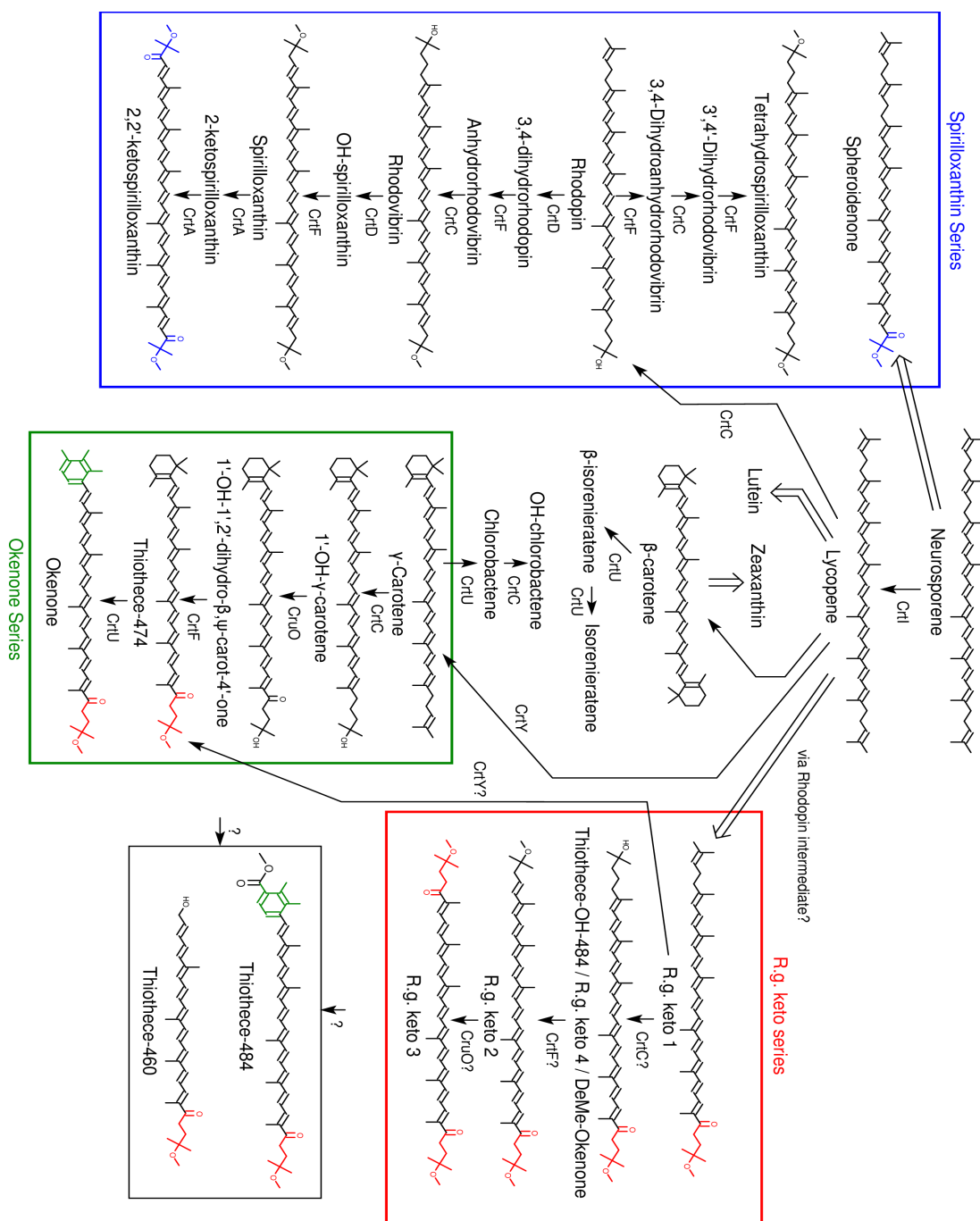


Figure 3.4: Biosynthetic pathways of carotenoids produced in purple sulfur bacteria, along with pathways to selected other carotenoids previously identified in meromictic lakes. Thick arrows indicate multiple steps and question marks after enzymes indicate predictions without experimental support. All 4' keto  $\psi$  ends are highlighted in red, 2' keto  $\psi$  ends in blue, and  $\chi$ -rings diagnostic of okenone are highlighted in green.



A third—albeit speculative—alternative for the production of Thiothece-484 is that it is under biological control and it, rather than okenone, is the terminal product of the okenone pathway. *Candidatus* ‘Thiodictyon syntrophicum’ strain Cad16<sup>T</sup> contains a ketolase (CruS), which has been shown *in vitro* to both add at the 2' position using molecular oxygen and simultaneously desaturate at the 3',4' position of 1'-hydroxyneurosporene and 1,1'-dihydroxyneurosporene (Vogl and Bryant, 2011). As *Candidatus* ‘Thiodictyon syntrophicum’ strain Cad16<sup>T</sup> contains the cruS gene and yet doesn't produce either of these neurosporene compounds, its *in vivo* function is unknown. Speculatively, it may be involved in oxidation of okenone to Thiothece-484 under high oxygen fugacities. Consistent with this theory, the cruS gene is also present in *Thiocapsa marina* (Vogl and Bryant, 2011), a known producer of Thiothece-484 (Caumette et al., 1985). Given this hypothesis, mass spectral analysis of carotenoids produced by *Candidatus* ‘Thiodictyon syntrophicum’ strain Cad16<sup>T</sup> should either confirm or deny the presence of Thiothece-484 and here would suggest that *Chromatiaceae* strain ML1 also actively produces Thiothece-484.

Similar late-stage modification processes are seen in several other anoxygenic phototrophs. The green sulfur bacterium *Chloroflexus aurantiacus* produces a greater proportion of glucosidic carotenoids under high light conditions (Schmidt et al., 1980). Purple non-sulfur bacteria are able to ketolate spirilloxanthin and spheroidene-series carotenoids in the 2' position (Pinta et al., 2003; Gerjets et al., 2009), and the purple non-sulfur bacterium *Rubrivivax gelatinosus* produces proportionally more keto-spirilloxanthins under semi-aerobic conditions versus more non-ketolated spirilloxanthin under anaerobic conditions (Takaichi and Shimada, 1999). These interpretations would be consistent with the greater abundance of Thiothece-484 observed in the

atmosphere-exposed shoreline sediment of Mahoney Lake, and the low abundance or absence of this compound in the 7-m chemocline.

In contrast to arguments about the biogenicity of Thiothece-484, it is probable that the identification of Thiothece-460 and the other minor carotenoids only observed in the Bligh-Dyer extracts are oxidative degradation products of okenone produced during that extraction. It is unclear why okenone in the sediment and shoreline samples was not similarly converted to Thiothece-460, as one might expect it to be formed among samples that had been extracted and handled identically. Thiothece-460 has previously been identified (Andrewes and Liaaen-Jensen, 1972) as a product from cultures of *Thiocystis gelatinosa* (formerly *Thiothece gelatinosa*; Pfennig and Trüper, 1971), although it may have had a similar genesis in those experiments.

#### *Sources and biosynthesis of okenone – challenges for future research*

Brocks and Schaeffer (2008) hypothesized that okenone is produced only in planktonic environments and therefore that it is a specific tracer for sulfide at shallow depths in open marine photic zones. They did, however, acknowledge that 3 of the 11 known okenone-producing species were isolated from benthic environments. In particular, this includes the mat-dwelling species *Thiohalocapsa halophila* (Caumette et al., 1991), the closest relative to the planktonic *Chromatiaceae* strain ML1. To interpret this finding, Brocks and Schaeffer (2008) suggested that okenone production did not occur when PSB species were living in benthic mats, as okenone had not been measured in any such non-planktonic systems at that time. Recently, however, okenone

was found in a microbial mat overlain by an oxic water column in Fayetteville Green Lake (Meyer et al., 2011). The presence of okenone was thus strongly suspected to be autochthonous, but there remained the possibility that lateral rafting of planktonic material from deeper layers in the lake also could have transported it to those mats. Similarly, lateral transport of PSB material is a known phenomenon in Mahoney Lake (Overmann et al., 1996b, and Chapter 1, this thesis). It therefore remains uncertain whether synthesis of okenone actually occurs outside of planktonic communities. However, given our understanding of Thiothece-484 as a closely related biosynthetic intermediate or terminal product of the okenone pathway, its presence in the shoreline sample but not in the 7-m water column sample may indicate that okenone is produced *in situ* in the shoreline. Alternatively, it could be interpreted strictly as a degradation product of okenone deposited in the littoral and basinal sediments by rafted planktonic PSB. Resolving these possibilities remains a challenge for future work, likely involving measurement of carbon and/or hydrogen isotope ratios of planktonic, shoreline, and basinal carotenoids, as well as further elucidation of the complete okenone biosynthetic pathway (Vogl and Bryant, 2011; 2012).

Further characteristics of okenone potentially complicate its use as a paleotaxonomic and paleoenvironmental tracer. Two moieties distinguish okenone from other carotenoids: the ketone in the 4' position and the  $\chi$ -ring (Figure 3.4). A 4' ketone also is found in the acyclic *R.g.* series of carotenoids produced by purple non-sulfur bacteria (Schmidt and Liaaen-Jensen, 1973). Similarly, recent work has shown that the  $\chi$ -ring also is present in some cyanobacterial carotenoids, e.g., synechoxanthin and renierapurpurin (Graham and Bryant, 2008), and as such, neither of the characteristic okenone moieties is unique to PSB. However, the combination of

both features is still believed to be confined taxonomically to the Chromatiaceae, but given uncertainty about what function okenone serves and its biosynthetic pathway, other possibilities should be examined. In particular, Thiothece-484 has a carbon backbone identical to okenone, so it may be an additional major precursor of okenane during diagenesis. This information will not substantially change geologic interpretations of okenane, since both Thiothece-484 and okenone likely are produced in the same pathway and by the same species, but caution is warranted as our understanding of okenone production is still incomplete.

Clearly a more in-depth understanding of the biosynthetic pathway of okenone is essential. Initially, Schmidt and Liaaen-Jensen (1973) proposed a pathway to okenone in which complete modification of the  $\psi$  end happened before the formation of the  $\chi$ -ring; this work was based on carotenoid extracts of purple non-sulfur bacteria (pathway through red box, Figure 3.4). In contrast, Vogl and Bryant (2011, 2012) proposed a scheme where cyclization to  $\gamma$ -carotene occurred first, followed by modification of the  $\psi$  end, and then the final rearrangement of the  $\beta$  ring to a  $\chi$  ring; this work was based on the individual expression of Chromatiaceae genes in *E. coli* (pathway through green box, Figure 3.4). Additionally, in both the spirilloxanthin pathway (blue box, Figure 3.4; Takaichi 2009) and in the pathway of Schmidt and Jensen (1973), it is proposed that ketolation takes place after methylation of the  $\psi$  end, while in the scheme of Vogl and Bryan (2011), ketolation precedes methylation. These two schemes are not necessarily incompatible, as purple non-sulfur bacteria (e.g. in Schmidt and Jensen's scheme) and purple sulfur bacteria (e.g. in Vogl and Bryant's scheme) may not share the same order of reactions in productions of their carotenoids. Understanding the genes, enzymes, and order of expression will

however be essential to further analysis of whether the 4' keto group in okenone biosynthesis is unique to PSB, or if it is homologous to the 4' keto group in the *R.g.* keto series of purple non-sulfur bacteria.

## **Conclusion**

This work confirms that the major Mahoney Lake pigments are okenone and bacteriochlorophyll *a*, as suggested previously, but there are also several interesting pigments present in the system, particularly Thiothece-484. It appears that Thiothece-484 is an abundant carotenoid in a subset of okenone-producing bacteria and that it is also found in geographically-distributed microbial mat communities. As our understanding of okenone production and diagenesis is incomplete, further examination of the intermediates in the biosynthetic pathways of the okenone and *R.g.* keto carotenoid series is warranted. Similarly, it will be important to achieve a better understanding of the functional properties of these two classes of purple carotenoids and the reason for the presence of the  $\chi$  ring only in PSB, but the 4'-keto group in both PSB and purple non-sulfur bacteria.

## References

- Airs RL, Atkinson JE, Keely BJ (2001) Development and application of a high resolution liquid chromatographic method for the analysis of complex pigment distributions. *Journal of Chromatography A*, **917**(1–2), 167–177.
- Andrewes AG, Liaaen-Jensen S (1972) Bacterial carotenoids. XXXVII. Carotenoids of thiorhodaceae 9. Structural elucidation of five minor carotenoids from *Thiothece gelatinosa*. *Acta Chemica Scandinavica*, **26**(6).
- Barua AB, Olson JA (1998) Reversed-phase gradient high-performance liquid chromatographic procedure for simultaneous analysis of very polar to nonpolar retinoids, carotenoids and tocopherols in animal and plant samples. *Journal of Chromatography B: Biomedical Sciences and Applications*, **707**(1-2), 69–79.
- Bialek-Bylka GE, Fujii R, Chen C-H, Oh-oka H, Kamiesu A, Satoh K, Koike H, Koyama Y (1998) 15-Cis-carotenoids found in the reaction center of a green sulfur bacterium *Chlorobium tepidum* and in the Photosystem I reaction center of a cyanobacterium *Synechococcus vulcanus*. *Photosynthesis Research*, **58**(2), 135–142.
- Bligh EG, Dyer WJ (1959) A rapid method of total lipid extraction and purification. *Canadian Journal of Physiology and Pharmacology*, **37**, 911–917.
- Bosshard PP, Santini Y, Grütter D, Stettler R, Bachofen R (2000) Bacterial diversity and community composition in the chemocline of the meromictic alpine Lake Cadagno as revealed by 16S rDNA analysis. *FEMS Microbiology Ecology*, **31**(2), 173–182.
- Britton G, Liaaen-Jensen S, Pfander HP (2004) *Carotenoids: Handbook*. Springer, Basel.
- Brocks JJ, Love GD, Summons RE, Knoll AH, Logan GA, Bowden SA (2005) Biomarker evidence for green and purple sulphur bacteria in a stratified Palaeoproterozoic sea. *Nature*, **437**(7060), 866–870.
- Brocks JJ, Schaeffer P (2008) Okenane, a biomarker for purple sulfur bacteria (Chromatiaceae), and other new carotenoid derivatives from the 1640 Ma Barney Creek Formation. *Geochimica et Cosmochimica Acta*, **72**(5), 1396–1414.
- Caumette P, Baulaigue R, Matheron R (1991) *Thiocapsa halophila* sp. nov., a new halophilic phototrophic purple sulfur bacterium. *Archives of Microbiology*, **155**(2), 170–176.
- Caumette P, Guyoneaud R, Imhoff JF, Süling J, Gorlenko V (2004) *Thiocapsa marina* sp. nov., a novel, okenone-containing, purple sulfur bacterium isolated from brackish coastal and marine environments. *International Journal of Systematic and Evolutionary Microbiology*, **54**(4), 1031–1036.
- Caumette P, Schmidt K, Biebl H, Pfennig N (1985) Characterization of a *Thiocapsa*

Strain Containing Okenone as Major Carotenoid. *Systematic and Applied Microbiology*, **6**(2), 132–136.

Coolen MJL, Overmann J (1998) Analysis of Subfossil Molecular Remains of Purple Sulfur Bacteria in a Lake Sediment. *Applied and Environmental Microbiology*, **64**(11), 4513–4521.

Coolen MJL, Overmann J (2007) 217 000-year-old DNA sequences of green sulfur bacteria in Mediterranean sapropels and their implications for the reconstruction of the paleoenvironment. *Environmental Microbiology*, **9**(1), 238–249.

Crowe SA, Jones C, Katsev S, Magen C, O'Neill AH, Sturm A, Canfield DE, Haffner GD, Mucci A, Sundby B, Fowle DA (2008) Photoferrotrophs thrive in an Archean Ocean analogue. *Proceedings of the National Academy of Sciences*, **105**(41), 15938–15943.

Davidson E, Cogdell RJ (1981) Reconstitution of carotenoids into the light-harvesting pigment-protein complex from the carotenoidless mutant of *Rhodospseudomonas sphaeroides* R26. *Biochimica et Biophysica Acta (BBA) - Bioenergetics*, **635**(2), 295–303.

Findenegg I (1937) Holomiktische und meromiktische Seen. *Internationale Revue der gesamten Hydrobiologie und Hydrographie*, **35**(1-6), 586–610.

Gerjets T, Steiger S, Sandmann G (2009) Catalytic properties of the expressed acyclic carotenoid 2-ketolases from *Rhodobacter capsulatus* and *Rubrivivax gelatinosus*. *Biochimica et Biophysica Acta (BBA) - Molecular and Cell Biology of Lipids*, **1791**(2), 125–131.

Graham JE, Bryant DA (2008) The Biosynthetic Pathway for Synechoxanthin, an Aromatic Carotenoid Synthesized by the Euryhaline, Unicellular Cyanobacterium *Synechococcus* sp. Strain PCC 7002. *Journal of Bacteriology*, **190**(24), 7966–7974.

Grice K, Cao C, Love GD, Twitchett RJ, Bottcher M, Grosjean E, Summons RE, Turgeon SC, Dunning W, Jin Y (2005) Photic Zone Euxinia During the Permian-Triassic Superanoxic Event. *Science*, **307**(5710), 706–709.

Hall Ken J, Northcote TG (2012) Meromictic Lakes. In: *Encyclopedia of Lakes and Reservoirs* (eds. Bengtsson L, Herschy RW, Fairbridge RW). Springer Netherlands, pp. 519–524.

Hall K J, Northcote TG (1990) Production and decomposition processes in a saline meromictic lake. *Hydrobiologia*, **197**(1), 115–128.

Hamilton TL, Bovee RJ, Sattin SR, Mohr W, Schaperdorth I, Gilhooly WP, Lyons TW, Pearson A, Macalady JL (2013) Sulfur dynamo in the phototrophic plate of a meromictic lake. *Geobiology*, (submitted).

Hebting Y (2006) Biomarker Evidence for a Major Preservation Pathway of Sedimentary Organic Carbon. *Science*, **312**(5780), 1627–1631.

Hodgson DA, Vyverman W, Verleyen E, Sabbe K, Leavitt PR, Taton A, Squier AH, Keely BJ (2004) Environmental factors influencing the pigment composition of in situ benthic microbial communities in east Antarctic lakes. *Aquatic microbial ecology*, **37**(3), 247–263.

Imhoff JF (2004) Taxonomy and physiology of phototrophic purple bacteria and green sulfur bacteria. In: *Anoxygenic Photosynthetic Bacteria* (eds. Blankenship RE, Madigan MT, Bauer CE). Kluwer Academic Publishers, Dordrecht, pp. 1–15.

Imhoff JF (2006) The Chromatiaceae. In: *The Prokaryotes* (eds. Dworkin M, Falkow S, Rosenberg E, Schleifer K-H, Stackebrandt E). Springer, New York, pp. 846–873.

Jenkyns HC (1980) Cretaceous anoxic events: from continents to oceans. *Journal of the Geological Society*, **137**(2), 171–188.

Johnston DT, Wolfe-Simon F, Pearson A, Knoll AH (2009) Anoxygenic photosynthesis modulated Proterozoic oxygen and sustained Earth's middle age. *Proceedings of the National Academy of Sciences*, **106**(40), 16925–16929.

Jørgensen BB, Fossing H, Wirsén CO, Jannasch HW (1991) Sulfide oxidation in the anoxic Black Sea chemocline. *Deep Sea Research Part A. Oceanographic Research Papers*, **38**, **Supplement 2**, S1083–S1103.

Koopmans MP, Köster J, Van Kaam-Peters HME, Kenig F, Schouten S, Hartgers WA, de Leeuw JW, Sinninghe Damsté JS (1996) Diagenetic and catagenetic products of isorenieratene: Molecular indicators for photic zone anoxia. *Geochimica et Cosmochimica Acta*, **60**(22), 4467–4496.

Leavitt PR (1993) A review of factors that regulate carotenoid and chlorophyll deposition and fossil pigment abundance. *Journal of Paleolimnology*, **9**(2), 109–127.

Liaaen-Jensen S (2004) Basic carotenoid chemistry. In: *Carotenoids in Health and Disease* (eds. Krinsky NT, Mayne ST, Sies H). Marcel Dekker, New York, pp. 1–30.

Liaaen-Jensen S, Lutnæs BF (2008) E/Z isomers and isomerization. In: *Carotenoids* Springer, pp. 15–36.

Lowe DJ, Green JD, Northcote TG, Hall KJ (1997) Holocene Fluctuations of a Meromictic Lake in Southern British Columbia. *Quaternary Research*, **48**(1), 100–113.

Lüthy L, Fritz M, Bachofen R (2000) In Situ Determination of Sulfide Turnover Rates in a Meromictic Alpine Lake. *Applied and Environmental Microbiology*, **66**(2), 712–717.

Lyons TW, Anbar AD, Severmann S, Scott C, Gill BC (2009) Tracking Euxinia in the Ancient Ocean: A Multiproxy Perspective and Proterozoic Case Study. *Annual Review of Earth and Planetary Sciences*, **37**, 507–534.

Meyer KM, Kump LR (2008) Oceanic Euxinia in Earth History: Causes and



Consequences. *Annual Review of Earth and Planetary Sciences*, **36**(1), 251–288.

Meyer KM, Macalady JL, Fulton JM, Kump LR, Schaperdoth I, Freeman KH (2011) Carotenoid biomarkers as an imperfect reflection of the anoxygenic phototrophic community in meromictic Fayetteville Green Lake. *Geobiology*, **9**(4), 321–329.

Montesinos E, Guerrero R, Abella C, Esteve I (1983) Ecology and Physiology of the Competition for Light Between *Chlorobium limicola* and *Chlorobium phaeobacteroides* in Natural Habitats. *Applied and Environmental Microbiology*, **46**(5), 1007–1016.

Nishihara M, Koga Y (1987) Extraction and Composition of Polar Lipids from the Archaeobacterium, *Methanobacterium thermoautotrophicum*: Effective Extraction of Tetraether Lipids by an Acidified Solvent. *Journal of Biochemistry*, **101**(4), 997–1005.

Northcote TG, Hall KJ (1983) Limnological contrasts and anomalies in two adjacent saline lakes. *Hydrobiologia*, **105**(1), 179–194.

Northcote TG, Halsey TG (1969) Seasonal Changes in the Limnology of Some Meromictic Lakes in Southern British Columbia. *Journal of the Fisheries Research Board of Canada*, **26**(7), 1763–1787.

Ohkouchi N, Nakajima Y, Okada H, Ogawa NO, Suga H, Oguri K, Kitazato H (2005) Biogeochemical processes in the saline meromictic Lake Kaiike, Japan: implications from molecular isotopic evidences of photosynthetic pigments. *Environmental Microbiology*, **7**(7), 1009–1016.

Overmann J, Beatty JT, Hall KJ, Pfennig N, Northcote TG (1991) Characterization of a dense, purple sulfur bacterial layer in a meromictic salt lake. *Limnology and Oceanography*, **36**(3), 846–859.

Overmann J, Fischer U, Pfennig N (1992) A new purple sulfur bacterium from saline littoral sediments, *Thiorhodovibrio winogradskyi* gen. nov. and sp. nov. *Archives of Microbiology*, **157**(4), 329–335.

Overmann J, Sandmann G, Hall KJ, Northcote TG (1993) Fossil carotenoids and paleolimnology of meromictic Mahoney Lake, British Columbia, Canada. *Aquatic Sciences*, **55**(1), 31–39.

Overmann J, Beatty JT, Krouse HR, Hall KJ (1996a) The sulfur cycle in the chemocline of a meromictic salt lake. *Limnology and oceanography*, **41**(1), 147–156.

Overmann J, Beatty JT, Hall KJ (1996b) Purple sulfur bacteria control the growth of aerobic heterotrophic bacterioplankton in a meromictic salt lake. *Applied and Environmental Microbiology*, **62**(9), 3251–3258.

Overmann J (1997) Mahoney Lake: a case study of the ecological significance of phototrophic sulfur bacteria. In: *Advances in Microbial Ecology 15* (ed. Jones JG). Springer

Science, New York, pp. 251–284.

Pfennig N, Markham MC, Liaaen-Jensen S (1968) Carotenoids of Thiorhodaceae. *Archiv für Mikrobiologie*, **62**(2), 178–191.

Pfennig N, Trüper HG (1971) New Nomenclatural Combinations in the Phototrophic Sulfur Bacteria. *International Journal of Systematic Bacteriology*, **21**(1), 11–14.

Pinta V, Ouchane S, Picaud M, Takaichi S, Astier C, Reiss-Husson F (2003) Characterization of unusual hydroxy- and ketocarotenoids in *Rubrivivax gelatinosus*: involvement of enzyme CrtF or CrtA. *Archives of Microbiology*, **179**(5), 354–362.

Repeta DJ (1993) A high resolution historical record of Holocene anoxygenic primary production in the Black Sea. *Geochimica et Cosmochimica Acta*, **57**(17), 4337–4342.

Romero-Viana L, Keely BJ, Camacho A, Vicente E, Rosa Miracle M (2009) Photoautotrophic community changes in Lagunillo del Tejo (Spain) in response to lake level fluctuation: Two centuries of sedimentary pigment records. *Organic Geochemistry*, **40**(3), 376–386.

Sanger JE (1988) Fossil pigments in paleoecology and paleolimnology. *Palaeogeography, Palaeoclimatology, Palaeoecology*, **62**(1–4), 343–359.

Schaeffer P, Adam P, Wehrung P, Albrecht P (1997) Novel aromatic carotenoid derivatives from sulfur photosynthetic bacteria in sediments. *Tetrahedron Letters*, **38**(48), 8413–8416.

Schmidt KARIN, Liaaen-Jensen S (1973) Bacterial carotenoids XLII: New Keto-carotenoids from *Rhodopseudomonas globiformis* (Rhodospirillaceae). *Acta Chem. Scand*, **27**(8), 3040–3052.

Schmidt K., Maarzahl M, Mayer F (1980) Development and pigmentation of chlorosomes in *Chloroflexus aurantiacus* strain Ok-70-fl. *Archives of Microbiology*, **127**(2), 87–97.

Senge MO, Smith KM, Blankenship RE, Madigan MT, Bauer CE (2004) Biosynthesis and structures of the bacteriochlorophylls. In: *Anoxygenic Photosynthetic Bacteria* Kluwer Academic Publishers, Dordrecht, pp. 137–151.

Sinninghe Damsté JS, Hopmans EC (2008) Does fossil pigment and DNA data from Mediterranean sediments invalidate the use of green sulfur bacterial pigments and their diagenetic derivatives as proxies for the assessment of past photic zone euxinia? *Environmental Microbiology*, **10**(6), 1392–1399.

Smith D, Scott J, Steele A, Cody G, Ohara S, Fogel M (2014) Effects of Metabolism and Physiology on the Production of Okenone and Bacteriochlorophyll a in Purple Sulfur Bacteria. *Geomicrobiology Journal*, **31**(2), 128–137.

Stal LJ, van Gernerden H, Krumbein WE (1985) Structure and development of a benthic marine microbial mat. *FEMS Microbiology Letters*, **31**(2), 111–125.

Steenbergen CLM, Korthals HJ, Dobrynin EG (1994) Algal and bacterial pigments in non-laminated lacustrine sediment: studies of their sedimentation, degradation and stratigraphy. *FEMS microbiology ecology*, **13**(4), 335–351.

Storelli N, Peduzzi S, Saad MM, Frigaard N-U, Perret X, Tonolla M (2013) CO<sub>2</sub> assimilation in the chemocline of Lake Cadagno is dominated by a few types of phototrophic purple sulfur bacteria. *FEMS Microbiology Ecology*, **84**(2), 421–432.

Summons RE, Powell TG (1986) Chlorobiaceae in Palaeozoic seas revealed by biological markers, isotopes and geology. *Nature*, **319**(6056), 763–765.

Takaichi S (2009) Distribution and Biosynthesis of Carotenoids. In: *The Purple Phototrophic Bacteria* (eds. Hunter CN, Daldal F, Thurnauer MC, Beatty JT). Springer Science, Dordrecht, pp. 97–117.

Takaichi Shinichi, Shimada K (1999) Pigment Composition of Two Pigment-Protein Complexes Derived from Anaerobically and Semi-Aerobically Grown *Rubrivivax gelatinosus*, and Identification of a New Keto-Carotenoid, 2-Ketospirilloxanthin. *Plant and Cell Physiology*, **40**(6), 613–617.

Tank M, Thiel V, Imhoff JF (2009) Phylogenetic relationship of phototrophic purple sulfur bacteria according to pufL and pufM genes. *International Microbiology*, **12**(3), 175–185.

Tonolla M, Peduzzi R, Hahn D (2005) Long-Term Population Dynamics of Phototrophic Sulfur Bacteria in the Chemocline of Lake Cadagno, Switzerland. *Applied and Environmental Microbiology*, **71**(7), 3544–3550.

Torsi G, Chiavari G, Laghi C, Asmundsdottir AM (1990) Responses of different UV-visible detectors in high-performance liquid chromatographic measurements when the absolute number of moles of an analyte is measured. *Journal of Chromatography A*, **518**, 135–140.

Villanueva J, Grimalt JO, de Wit R, Keely BJ, Maxwell JR (1994) Chlorophyll and carotenoid pigments in solar saltern microbial mats. *Geochimica et Cosmochimica Acta*, **58**(21), 4703–4715.

Vogl K, Bryant DA (2011) Elucidation of the Biosynthetic Pathway for Okenone in *Thiodictyon* sp. CAD16 Leads to the Discovery of Two Novel Carotene Ketolases. *Journal of Biological Chemistry*, **286**(44), 38521–38532.

Vogl K, Bryant DA (2012) Biosynthesis of the biomarker okenone:  $\chi$ -ring formation. *Geobiology*, **10**(3), 205–215.

Wetzel RG (2001) *Limnology: Lake and River Ecosystems*. Elsevier, San Diego.

Wyman C, Sloan P-P, Shirley P (2013) Simple Analytic Approximations to the CIE XYZ Color Matching Functions. *Journal of Computer Graphics Techniques*, **2**(2), 1–11.

Xiong J, Fischer WM, Inoue K, Nakahara M, Bauer CE (2000) Molecular Evidence for the Early Evolution of Photosynthesis. *Science*, **289**(5485), 1724–1730.

Züllig DH, Rheineck SG (1985) Pigmente phototropher Bakterien in Seesedimenten und ihre Bedeutung für die Seenforschung. *Swiss Journal of Hydrology*, **47**(2), 87–126.

## CHAPTER 4

### Identifying Associated Genes in the Environment through Tetranucleotide Distances

#### Introduction

##### *Justification*

With the rise of high-throughput DNA sequencing, it has become possible to find and compare the total complement of metabolic pathways present in environmental metagenomes (Dinsdale et al., 2008; Tringe et al., 2005). Such approaches provide a broad picture of an environmental community, but equally important is to understand how these functional pathways are distributed among and co-located in individual organisms or taxa. Our understanding of geological, biogeochemical, and microbial processes would be enhanced by such a deeper understanding. Are microbes recruited into or maintained in communities based on the biological functions they perform (Burke et al., 2011)? What do specific biosynthetic products tell us about the metabolism and environment of their source organisms (Brocks and Pearson, 2005)?

The concept of linking pathways in metagenomes is a generalization of the familiar approach of linking specific phylogenetic identities to specific enzymatically-driven processes (Gray and Head, 2001; Suenaga, 2012). For example, the question of which microbes degrade plant matter in cow rumens (Ferrer et al., 2005; Hess et al., 2011) is equivalent to asking which 16S ribosomal RNA sequences co-occur with known plant-matter degradation pathways? The question can also be reversed to find metabolic functions associated with specific clades: for example, what ecological functions do Planctomycetes perform in marine upwelling systems?

Woebken et al. (2007) screened fosmids containing 16S ribosomal RNA genes of Planctomycetes and found genes for metabolism of single-carbon substrates, suggesting a role at the base of the carbon cycle.

Such concepts may be useful for linking metabolism (and hence environmental context) directly to lipid biosynthesis, which is a major goal of organic geochemistry. For example, isorenieratene is a terminal product of green sulfur bacteria, the Chlorobiaceae, in sulfidic marine ecosystems (e.g., Grice et al., 2005). Chlorobiaceae are photoautotrophs that use sulfide as an electron donor, so isorenieratene is believed to be a good proxy for photic zone euxinia. However, the aerobic bacterium *Brevibacterium linens* also produces isorenieratene (Kohl et al., 1983). It has a closely-related sister species, *Brevibacterium oceanii*, that lives in marine systems (Bhadra et al., 2008), but it is not currently known if *B. oceanii* has the capacity to produce isorenieratene. Although sequencing *B. oceanii* and performing a lipid analysis would answer both the genetic and functional aspects of this question, the result could not be generalized to other species. A more general method is needed to resolve problems like these, such cases in rarer, unculturable organisms.

The ecological theory of a “core microbiome” is that in similar environments, certain core microbial members tend to occupy central roles (Shade and Handelsman, 2012). For example, deep waters of the North Atlantic and hydrothermal vents are dominated by a few members of the Proteobacteria, but thousands of other species can be identified at much lower abundances (Sogin et al., 2006). Similar results have been reported from the Sea of Marmara where species

of Proteobacteria, Planctomycetes and Bacteroidetes were most abundant, but many other species contributed to the ecosystem diversity (Quaiser et al., 2010). Despite their rarity, these species may perform important metabolic roles in the community: *Desulfosporosinus* sp. are sulfate-reducers which accounted for only 0.006% of the total microbial community in a peatland, but nonetheless were found to play a quantitative role in sulfate reduction (Pester et al., 2010). Similarly, the methylophilic *Methylophilum mobilis* comprised less than 0.4% of total bacteria in a lake by 16S sequencing, but they play a major role in cycling single-carbon compounds (Kalyuzhnaya et al., 2008).

Are these rare community members specializing into niches and losing pathways to syntrophic partners? Raes and Bork (2008) proposed a concept of clusters of orthologous groups (COG) richness per genome equivalent to determine how many “functions” would be predicted for each organism in a community. Several recent studies (Tyson et al., 2004; Jones et al., 2011; Swingley et al., 2012) have examined these concepts in specific, abundant organisms to determine how they contribute to their overall microbial community. Understanding how pathways co-occur in other community members remains a current goal.

Here we approach this problem by assuming that genes with similar tetranucleotide abundance patterns (explained later) either come from the same organism or from organisms that are recently evolutionarily related, either because of common descent or lateral gene transfer (Perry and Beiko, 2010). If two genes, each of which is essential for a *different* metabolic pathway, have

similar tetranucleotide usage patterns, then it is likely that they are, or recently have been, present in the same organism.

### *Linking Genes*

The most common methods for finding gene associations in microbes are intensive sequencing and gene-orthology associations, in some cases in combination with cell separation. Technological developments now allow long sequences to be obtained from single cells (Raghunathan et al., 2005; Zhang et al., 1992), in some cases up to 70% of the original genome (Ishoe et al., 2008). Similarly, deep sequencing of complex metagenomes can lead to nearly complete assembly of minor species, if such a species has minimal strain-level or species-level diversity (Iverson et al., 2012). Such long contiguous sequences allow fairly complete descriptions of the species' metabolic complement. For example, single-cell amplification of an individual from the uncultivated TM7 phylum revealed a great deal of information, although key details of its carbon source and energy acquisition strategies remain unanswered (Podar et al., 2007). A limitation of all of these strategies is that it is difficult to address multiple species simultaneously.

The identification of physically-close gene cassettes or operons is a common technique for finding genes that participate in a single pathway. As a means of finding clusters of genes presumed to be involved in the same pathway, Overbeek et al. (1999) surveyed orthologous genes that occurred within 300 basepairs of each other in more than one genome. Their approach has been extended by several other authors to generate large databases of microbial gene clusters



(Mavromatis et al., 2009), link environmental contexts to these clusters (Bohnebeck et al., 2008), and improve binning accuracy of metagenomic sequences (Weng et al., 2010). These methods do not, however, detect associations of genes that are located more distantly on the chromosome or on different chromosomes, if present.

With advances in high-throughput sequencing and assembly, someday it may be possible to directly sequence every base from every microbe in an environment. Despite this, it is likely to remain a challenge to identify all co-occurring pathways. Some microbes have multiple chromosomes or several large plasmids (Suwanto and Kaplan, 1989), and thus complete assembly is not equivalent to cell-level genomic content. Additionally, many organisms promiscuously exchange genetic material, so a single genome (single cell) is only a portion of the “pan-genome” of that species (Medini et al., 2005). Although lateral gene transfer can occur across domains, most transformation and transduction is limited to similar species (Frost et al., 2005). A need remains for a technique to evaluate the metabolic potential of the full “pan-genome” of a given taxon.

### *Tetranucleotide Percentages*

Two genes can be assumed to come from a single organism if they are sequenced and assembled onto the same scaffold. If two given genes are not found on any of the same scaffolds, their association in one organism can still be determined using statistical properties of sequence data. Such approaches include binning by % GC content, but this distinguishes sequences in one dimension and is only useful for low-diversity samples (Tyson et al., 2004). Analogously,

different species use different abundances of tetranucleotide and other so-called k-mer patterns in their genomes (Karlin et al., 1994). These different abundances are specific to clades and can distinguish from the level of genus to, in some cases, subspecies (Pride et al., 2003; Sharon et al., 2012). These usage patterns have allowed the separation of different clades in complex metagenomic data (Dick et al., 2009; Teeling et al., 2004a).

Tetranucleotide percentages have found most of their practical uses in classifying unknown sequences into taxonomic bins for further assembly and annotation. Teeling et al. (2004b) used Pearson correlation coefficients between mean and standard deviation-adjusted tetranucleotide percentages of 116 known genomes to classify sequences from 6 metagenomes. The requirement for prior data, and the assumption that environmental metagenomes behave statistically similarly to known genomes, have rendered this technique less used. Recent approaches use clustering methods such as self-organizing maps (Abe et al., 2003; Dick et al., 2009; Uehara et al., 2011; Weber et al., 2011; Hamilton et al., in prep.). Self-organizing map techniques are very good at binning abundant species, but as diversity scales up they become extremely computationally expensive and are incapable of properly binning very poorly represented organisms. Often low-abundance bins are filtered out (Saeed et al., 2012), and as a result, important metabolisms in low-abundance organisms may be neglected.

Outside of their use in binning, tetranucleotide percentages still give a great deal of information about biological diversity in microbial ecosystems. Jiang et al. (2012) found that tetranucleotide frequencies combined with other k-mer frequencies from metagenomic samples gave insight into

diversity changes across those samples. Differences in the tetranucleotide usage patterns of the samples reflect the evolutionary trajectories of each microbe, including its nearest relatives and lateral-gene transfer partners (Perry and Beiko, 2010). Because tetranucleotide patterns tell us something about genetic exchange in addition to microbial identity, they can be useful for understanding the distribution not only of organisms, but also of pathways in these organisms.

Here we avoid the binning approach and instead use the tetranucleotide percentage of a given scaffold as the “address” of the genes on that scaffold. We then interpret the distances between these “addresses” as the strength of the linkage between the genes on those scaffolds. This is somewhat analogous to the study of ecological interactions, wherein the two-dimensional distance between species is used as a proxy for the strength of their ecological relationships (Andersen, 1992).

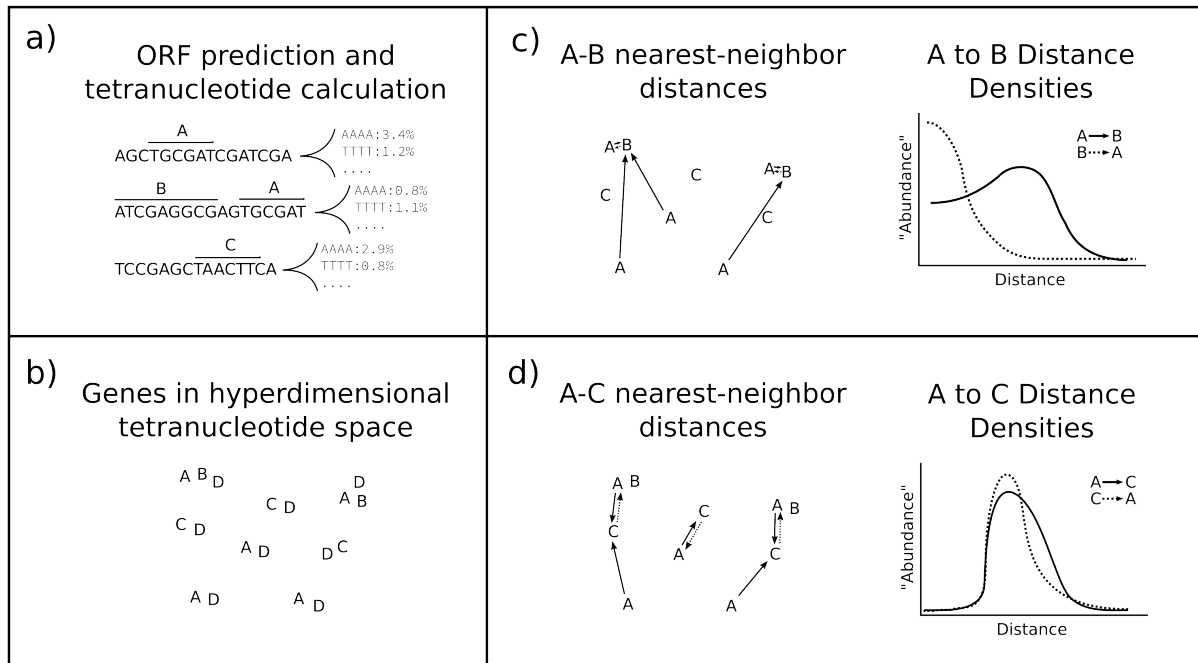


Figure 4.1: Illustration of how tetranucleotide score distances are calculated. See text for details.

## Approach

### *Hyperdimensional geometric tetranucleotide associations*

The general approach is as shown in Figure 4.1. Tetranucleotide frequencies are computed for assembled DNA contigs (Figure 4.1a), resulting in a histogram of frequency of occurrence of each of the 137 distinct tetranucleotides (Dick et al., 2009). This defines the “address” in 137-coordinate hyperdimensional space for all genes (open reading frames; ORFs) on that contig, here simplified to two dimensions for illustrative purposes (Figure 4.1b). In this cartoon, gene A represents a relatively common gene, likely to be found in many organisms in a population; gene B represents a rare gene that is part of a metabolic pathway dependent on A, so it cannot occur without A, although the reverse can occur. Gene C is another gene of modest frequency in the

community, but it is unassociated with either A or B. The Euclidian nearest-neighbor distances from one of these genes to another are thus asymmetric (Figure 4.1c): the  $A \rightarrow B$  distances show that some As are closely located in space to B's, while other As are not. Conversely, all Bs sit closely with As. The calculated abundance of geometric association lengths  $A \rightarrow B$  thus has significant magnitude both at distance zero (for As that have a B) and also at greater distances (for the unassociated As), while  $B \rightarrow A$  has the highest relative abundances only at short distances from A. Similarly,  $A \rightarrow C$  is bimodal, with some As quite close to a C, but with other As farther away (Figure 4.1d).

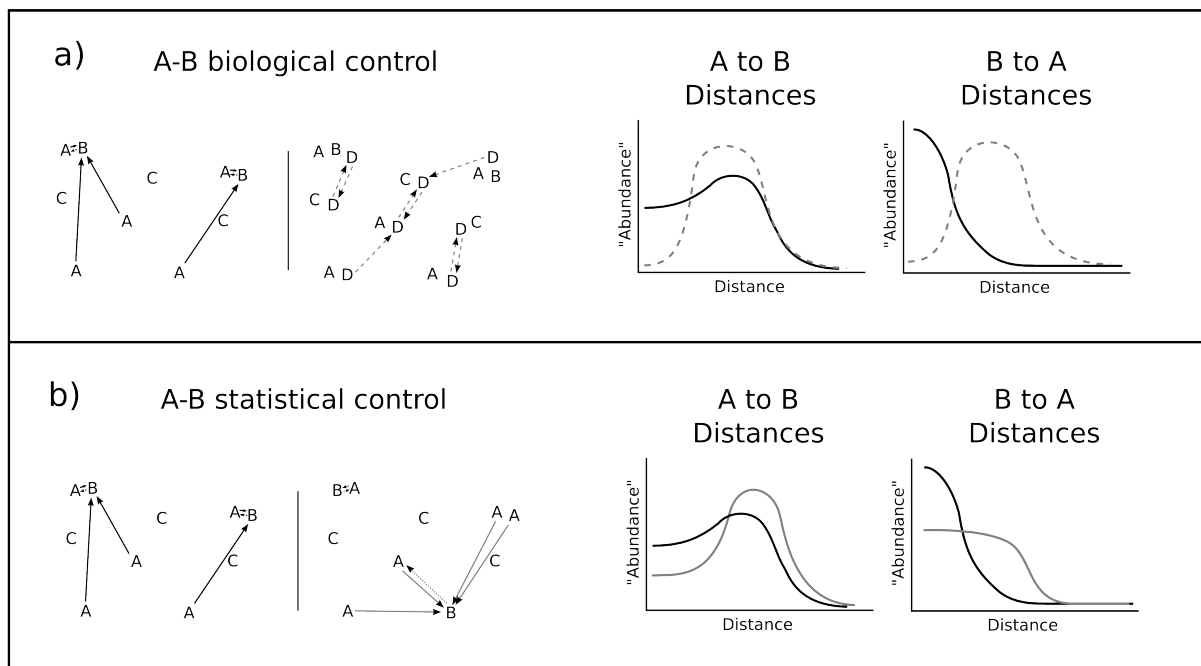


Figure 4.2: Illustration of controls for tetranucleotide association. See text for details.

Although such geometric distances are easy to compute, the goal is to determine if these distances reflect significant association of any two genes. This requires that each pair be

referenced to a control designed for neutrality of association (Figure 4.2). Such a control can be envisioned in two ways: (i) either the frame of reference is the distance between taxa in the community, or (ii) the frame of reference is a randomized distribution of the two genes being compared.

We refer to (i) as the biological control (Figure 4.2a). The simplest such control would be the distribution of a conserved gene, e.g., *rpoA* (here symbolized as gene D; Figure 4.2a). All taxa in a community must contain D, and the distances between them (symmetric;  $D \leftrightarrow D$ ) describe the phylogenetic “closeness” of a community. In the cartoon,  $A \rightarrow B$  associations are not always more closely related than are nearest-neighbor taxa ( $D \leftrightarrow D$ ) (Figure 4.2b); conversely, all  $B \rightarrow A$  distances are nearer than any two nearest-neighbor taxa. This indicates the  $B \rightarrow A$  gene pair is always associated within an organism and is never separated among neighboring organisms. However, the distance spectrum calculated in this control suffers from a data density bias. The frequency of occurrence of *rpoA* is a function of the abundance of taxa and the microdiversity of their genes. In the extreme case in which infinite data occupied one region of the total hyperdimensional space (e.g., an ultra-low-diversity community, dominated by many slightly-divergent strains of one species), all abundance vs. distance density plots would skew to the origin (Figure 4.2b), and all gray dashed (control) and black (gene-gene) lines would become identical.

An alternative approach is to apply a statistical randomization (option ii; Figure 4.2c). Our null hypothesis is that any two genes  $A \rightarrow B$  are distributed independently of each other. If this is true

then a random permutation of all data points classified as either A or B should not change the distance profile between them; this is equivalent to forcing  $A \rightarrow B$  and  $B \rightarrow A$  to be symmetrical. This is also similar to declaring that A and B are always hosted by different taxa. When the true data approach this statistical randomization, the genes can be viewed as “not associated”. The opposite case is when A is completely dependent on B. In this case, randomly reassigning the genes in the statistical control will increase the nearest-neighbor distance between them. The distance spectrum of truly associated genes will lie closer to the origin than the distance distribution of the control (Figure 4.2d). To execute this control, we randomly permute the positions of any two genes A and B among the union of both of their positions.

### *Gene Selection*

To apply this approach to natural samples, we select genes specific to known metabolic and biosynthetic pathways. Each gene must be essential to the metabolic pathway being investigated and should not be a strong homolog of other proteins of different function. For example, digeranylgeranylglycerol phosphate synthase (DGGGPS) and chlorophyll synthase (ChlG) are close homologs, but respectively they perform the second alkylation of archaeal ether lipids and the phytol esterification of chlorophyll. Avoiding such cases of sequence homology is even more challenging for core metabolisms than for biosynthetic pathways, since, for example, there are only subtle differences between many/most of the respiratory chain cytochromes. When possible, the genes encoding the target protein also should be present only once in a genome. In this work, all protein sequences were subjected to analysis by pBLAST and String to avoid cases such as the DGGGPS vs. ChlG example above.

As positive controls, we include genes known to co-occur in the same pathways. They are added as pairs of genes that always should exist together, either because they are part of a known operon or are both essential to a given pathway. The RNA polymerase genes *rpoA* and *rpoB* are part of the same operon and should occur in every bacterial species. Functional gene examples include *dsrA* and *dsrB* for sulfate reduction, and *psbA* and *psbB* for photosynthesis: one never should be present without the other. Finally, variants of the same gene from different species are also included when there are substantial sequence differences across a broad range of taxa; this helps ensure that more variants of a given gene are detected in the analysis.

## **Methods**

### *Metagenomes*

The approach was tested on metagenomes obtained from the 7 meter deep oxic-anoxic interface and 8 meter deep sulfide-rich layers of Mahoney Lake (49°17'N, 119°35'W) as described in Klepac-Ceraj et al. (2012). DNA was extracted and metagenomes were obtained as described in Hamilton et al. (in prep.). Briefly, DNA was extracted with an e.Z.N.A SP Plant Maxi Kit (Omega Biotek); fragmentation and library preparation were performed commercially the by North Carolina State University Genomic Sciences Laboratory; paired-end 150 bp Illumina (HiSeq 2500) sequencing was performed at the Harvard Center for Systems Biology. The raw data was quality filtered with Trimmomatic (Lohse et al., 2012) to remove the first three bases, Illumina TruSeq adapter sequences and any raw sequences of low quality. Only sequences with at least 50 basepairs in both the forward and reverse direction were retained. Scaffolds were



assembled from raw FASTQ genomic data by IDBA-UD using eight threads with default parameters (Peng et al., 2012). Open reading frames (ORFs) were identified from the assembled contigs using MetaGenemark with the supplied ORF prediction model (Zhu et al., 2010).

### *Gene Prediction*

Using the amino acid sequence of each of the genes listed in Table 4.1, a PSIBLAST search of the Genbank non-redundant (NR) database was conducted, returning all sequences with an E-value  $\leq 1e-10$ , a moderately stringent homology cutoff (Altschul et al., 1997). These sequences were then aligned using MUSCLE (Edgar, 2004) with 2 iterations (maxiters=2) as recommended by the MUSCLE documentation for balancing speed and accuracy in large alignments. The MUSCLE alignments were converted to Stockholm format for input into HMMBuild, a program included in HMMER 3.0 (Eddy, 2011), which used them to build hidden Markov models (HMM) for gene prediction.

Using the HMMER 3.0 program HMMScan with default settings, these models were compared to the ORF tables generated by MetaGenemark to identify putative gene homologs in the metagenomic dataset. The best match of a given metagenomic ORF was considered to be the identity of that ORF for further computation. For example, despite statistically significant matching of *shC*, *osc*, and *casI* to a single ORF, the ORF is defined only as the one of these with the lowest (best) E-value score. Additionally, to reduce false positives, only matches with E-values of less than  $10e-10$  were retained.

Table 4.1: List of genes used for finding gene associations. (\*) This enzyme also exists in acetyl-CoA autotrophs. (\*\*) Also used by a wide variety of acetate oxidizers.

Pathway	Gene	Enzyme Name	GI	Species
<b>Core metabolisms</b>				
Photosystem I, Cyanobacteria	psaA	Photosystem I P700 apoprotein A1	48058	Synechocystis sp. PCC 6803
	psaB	Photosystem I P700 apoprotein A2	48059	Synechocystis sp. PCC 6803
Photosystem II, Cyanobacteria	psbA	Photosystem II protein D1	16329178	Synechocystis sp. PCC 6803
	psbB	Photosystem II protein CP47	1001216	Synechocystis sp. PCC 6803
Photosystem I, Proteobacteria	pufM	Photosynthetic reaction center M	260161129	Allochromatium vinosum DSM 180
	pufL	Photosynthetic reaction center L	260161128	Allochromatium vinosum DSM 180
Photoheterotroph	pr	proteorhodopsin	149688766	Candidatus Pelagibacter ubique HTCC1002
Rubisco, form I	rbcL	Ribulose biphosphate carboxylase oxygenase	1710041	Synechocystis sp. 6803
Rubisco, form II	rbcL	Ribulose biphosphate carboxylase oxygenase	77389277	Rhodobacter sphaeroides 2.4.1
Phototrophic iron oxidation	pioA	Fe oxidase oxidoreductase CytC	119331447	Rhodopseudomonas palustris TIE-1
Phototrophic iron oxidation	pioC	Fe oxidase HiPIP	119331451	Rhodopseudomonas palustris TIE-1
N <sub>2</sub> fixation	nifH	nitrogenase subunit H	77464110	Rhodobacter sphaeroides 2.4.1
Chemoautotrophic iron oxidation	iro	Fe oxidase HiPIP	218666489	Acidithiobacillus ferrooxidans ATCC 23270
Aerobic iron/sulfur oxidation	coxB	aa3-type CytC oxidase	198284826	Acidithiobacillus ferrooxidans ATCC 53993
Manganese oxidation	ompC	multicopper oxidase, Mn-type	39984643	Geobacter sulfurreducens PCA
Autotrophic nitrification, Archaea	amoA	ammonia monooxygenase	166783464	Nitrosopumilus maritimus SCM1
Autotrophic nitrification, Bacteria	amoA	ammonia monooxygenase	3282845	Nitrosococcus oceanus ATCC 19707
Aerobic methane oxidation	mmoZ	particulate methane monooxygenase	53804672	Methylococcus capsulatus str. Bath
Photoautotrophic sulfide oxidation	sqr	sulfide-quinone oxidoreductase	77465561	Rhodobacter sphaeroides 2.4.1

*Table 4.1 (Continued): List of genes used for finding gene associations.*

Pathway	Gene	Enzyme Name	GI	Species
Thiosulfate oxidation, SOX cluster	sqr	sulfide-quinone oxidoreductase	288897230	Allochromatium vinosum DSM 180
	soxY		288897197	Allochromatium vinosum DSM 180
	soxZ		288897198	Allochromatium vinosum DSM 180
	soxA		288941882	Allochromatium vinosum DSM 180
	soxY		145558159	R. sphaeroides ATCC 17025
	soxZ		145558158	R. sphaeroides ATCC 17025
	soxA		145558157	R. sphaeroides ATCC 17025
Nitrate reduction	narG	respiratory nitrate reductase	75674970	Nitrobacter winogradskyi Nb-255
Nitrite reduction	nirK	Cu nitrite reductase	192292984	Rhodopseudomonas palustris TIE-1
Heterotroph, sulfate-reducing	dsrA	dissimilatory sulfite reductase A	219869855	Desulfovibrio desulfuricans ATCC 27774
	dsrB	dissimilatory sulfite reductase B	220905555	Desulfovibrio desulfuricans ATCC 27774
Methanogen, all	mcrA	methyl Co-M reductase A	20093330	Methanosarcina acetivorans C2A
Methanogen, CO <sub>2</sub> -reducing	frhB	F420-reducing dehydrogenase	20089855	Methanosarcina acetivorans C2A
Methanogen, aceticlastic*	cdhD	Corrinoid:H4SPT methyltransferase	20089887	Methanosarcina acetivorans C2A
Homoacetogenesis**	fdhA	NADP-dependent formate dehydrogenase	148283121	Moorella thermoacetica ATCC 39073
<b>Substrate access</b>				
Lignin oxidation (fungal)	lpg1	Lignin peroxidase	170200	Trametes versicolor
Cellulose degradation (fungal, bacterial)	cel	Endocellulase	392559170	Trametes versicolor FP-101664 SS1
	cel	Endocellulase	313204241	Paludibacter propionicigenes WB4
Cellulose degradation (bacterial)	bglA	Cellobiase (beta-glucosidase)	319952356	Cellulophaga algicola DSM 14237
Anaerobic degradation of aromatics	bamA	benzoyl-CoA hydrolase	418065883	G. metallireducens RCH3

Table 4.1 (Continued): List of genes used for finding gene associations.

Pathway	Gene	Enzyme Name	GI	Species
<b>Lipid synthetic pathway</b>				
MVA isoprenoids, bacterial and archaeal	mvk	mevalonate kinase	161527823	Nitrosopumilus maritimus SCM1
MEP isoprenoids, bacterial	dxr	deoxyxylulose phosphate reductase	288941764 <sup>180</sup>	Allochromatium vinosum DSM
Archaeal ether lipid	gggps	GG-glycerol phosphate synthase	160340604	Nitrosopumilus maritimus SCM1
Sulfolipid	sqdB	Sulfite:UDP-glucose sulfotransferase	1652231	Synechocystis sp. PCC 6803
Phospholipid	cdsA	phosphate cytidylyltransferase	77463259	Rhodobacter sphaeroides 2.4.1
	cdsA	phosphate cytidylyltransferase	288941763 <sup>180</sup>	Allochromatium vinosum DSM
	cdsA	phosphate cytidylyltransferase	400756568	Geobacter sulfurreducens PCA
	cdsA	phosphate cytidylyltransferase	16330180	Synechocystis sp. PCC 6803
Glycolipid, cyanobacterial	mgldD	DAG-glucosyltransferase MgldD	16330846	Synechocystis sp. PCC 6803
Glycolipid, Chlorobi	mgdA	DAG-galactosyltransferase MgdA	21673183	Chlorobium tepidum TLS
Betaine lipid	btaA	DAG aminocarboxypropyl transferase	77464435	Rhodobacter sphaeroides 2.4.1
Ornithine lipid	olsB	ornithine N-acyltransferase	77465008	Rhodobacter sphaeroides 2.4.1
Hopanoid	sqhC	squalene-hopene cyclase	53804820	Methylococcus capsulatus str. Bath
Steroid	osc	oxidosqualene-lanosterol cyclase	53803023	Methylococcus capsulatus str. Bath
	casI	oxidosqualene-cycloartenol cyclase	310822657	Stigmatella aurantiaca DW4/3-1
Carotenoid	crtI	phytoene desaturase, lycopene- forming	343801970 <sup>180</sup>	Allochromatium vinosum DSM
Carotenoid	crtP	phytoene desaturase, zeta- carotene-forming	16330439	Thiocapsa marina 5811
	crtC	hydrolase	2509050511	Thiocapsa marina 5811
	crtF	methylase	2509050504	Thiocapsa marina 5811
	cruO	ketolase		Thiocapsa marina 5811
	crtA	spirilloxanthin and/or spheroidene monooxygenase		Rubrivivax gelatinosus IL144
	cruE	carotene desaturase/methyltransferase	2509050500	Thiocapsa marina 5811
	crtU	beta-carotene desaturase	170077864	Synechococcus sp. PCC 7002

Table 4.1 (Continued): List of genes used for finding gene associations.

Pathway	Gene	Enzyme Name	GI	Species
Lactone/quorum sensing	crtY	lycopene cyclase	2509050499	Thiocapsa marina 5811, DSM 5653
	cruA	lycopene cyclase	146428530	Chlorobium phaeobacteroides
	crtL	lycopene beta cyclase	33865262	Synechococcus sp. WH 8102
	luxI	acyl-homoserine lactone synthase	30350260	Aliivibrio fischeri
	cerI	acyl-homoserine lactone synthase	77463695	Rhodobacter sphaeroides 2.4.1
	raiI	acyl-homoserine lactone synthase	254488731	Roseobacter sp. GAI101
Chlorophyll synthesis	por	Light-dep. protochlorophyllide oxidoreductase	16331782	Synechocystis sp. PCC 6803
Bacteriochlorophyll synthesis	bchF	2-vinyl bacteriochlorophyllide hydratase	288942344	Allochromatium vinosum DSM 180
<b>Control</b>				
RNA polymerase	rpoA	RNA polymerase subunit A	380879662	Thiorhodovibrio sp. 970
RNA polymerase	rpoB	RNA polymerase subunit B	380879630	Thiorhodovibrio sp. 970
RNA polymerase	rpoC	RNA polymerase subunit C	380879631	Thiorhodovibrio sp. 970

For each ORF identified as gene A, we calculated the Euclidian distance (n-th root of the sum of n squared distances) to the nearest neighboring ORF identified as gene B in tetranucleotide space. As the nearest neighbor distance is asymmetric between genes, the distance from gene B to gene A was also calculated. If only a few copies of a given gene are detected, there is a higher probability that random correlation will generate a false positive result, so we limited all A↔B comparisons to genes having more than 5 identified ORFs in the datasets. Finally, the controls (explained above) were calculated. The biological control was the distance from each *rpoA* gene to its nearest neighboring *rpoA* gene; this distance frequency is symmetrical. The statistical control between all pairs of A and B was calculated as a random permutation of the two genes between their two positions. To fully randomize this statistical control, the distance frequency spectrum was calculated from 500 bootstrap replicates for each pair.

### *Tetranucleotide Distance Calculation*

All steps were performed using a custom Python script (available online at <https://github.com/bovee/GeneLinkageScripts>) that used Numpy and Scipy for mathematical calculations, Screeed for speeding up scaffold lookup and Matplotlib for plotting. First, tetranucleotide percentages were computed for the full length of the contig associated with the ORF being analyzed. Corresponding reverse-complement tetranucleotides were combined as in Dick et al. (2009), and tetranucleotides containing scaffolding gaps were binned into the tetranucleotide NNNN for a total tetranucleotide dimensionality of 137.

To generate accurate histograms, tetranucleotide percentages must be calculated on sequences of at least 1000 basepairs, and longer sequences result in better classifications (Abe et al., 2003; McHardy and Rigoutsos, 2007). As a preliminary cutoff, we examined only scaffolds longer than 1000 basepairs. By culling smaller scaffolds, we likely remove low-coverage clades that are not abundant enough to assemble. This may remove some minor species that nonetheless are important biogeochemically or ecologically. Other clades that may be overlooked include those that do not assemble well because of highly repetitive genomes, high subspecies diversity, or many plasmids. Better assemblers may fix the first problem, but only higher sequencing depth will help solve the others. Regardless, such concerns are not unique to this work.

Although the distance results can be compared graphically (as in Figures 4.1, 4.2) to determine gene-gene relationships, statistical measures are needed to quantify the strength of the coupling. We used one-sided Mann-Whitney U probabilities between each set of gene-gene and corresponding gene-control arrays, as computed using the statistical programming language R using the *wilcox.test* function (R Development Core Team, 2008). The output is the probability that the underlying distribution of the  $A \rightarrow B$  gene-gene distances is larger than the distribution of the controls, i.e., that  $A \rightarrow B$  is farther away than would be generated by a random distribution of these genes. Therefore, a small  $p$ -value indicates that  $A \rightarrow B$  distances are either random, or closer together than random.. This is different from the two-sided Mann-Whitney U which is a non-parametric test of whether the two distributions are larger *or* smaller than each other. To calculate  $p$ , the observations for both the control distances and the  $A \rightarrow B$  distances are combined and ordered and the sum of the ranks for the control are used to make a U statistic. The distribution of the U statistic is known and can then be used to calculate a probability that the two distributions are not different. This is different from the two-sided Mann-Whitney U which is a non-parametric test of whether the two distributions are larger *or* smaller than each other. As we are not looking for the case where the sample distribution is larger than the control (e.g. if two genes “repulsed” each other and would never be found together), we use the one-sided Mann-Whitney, although that alternative case is still interesting and worthwhile of examination.

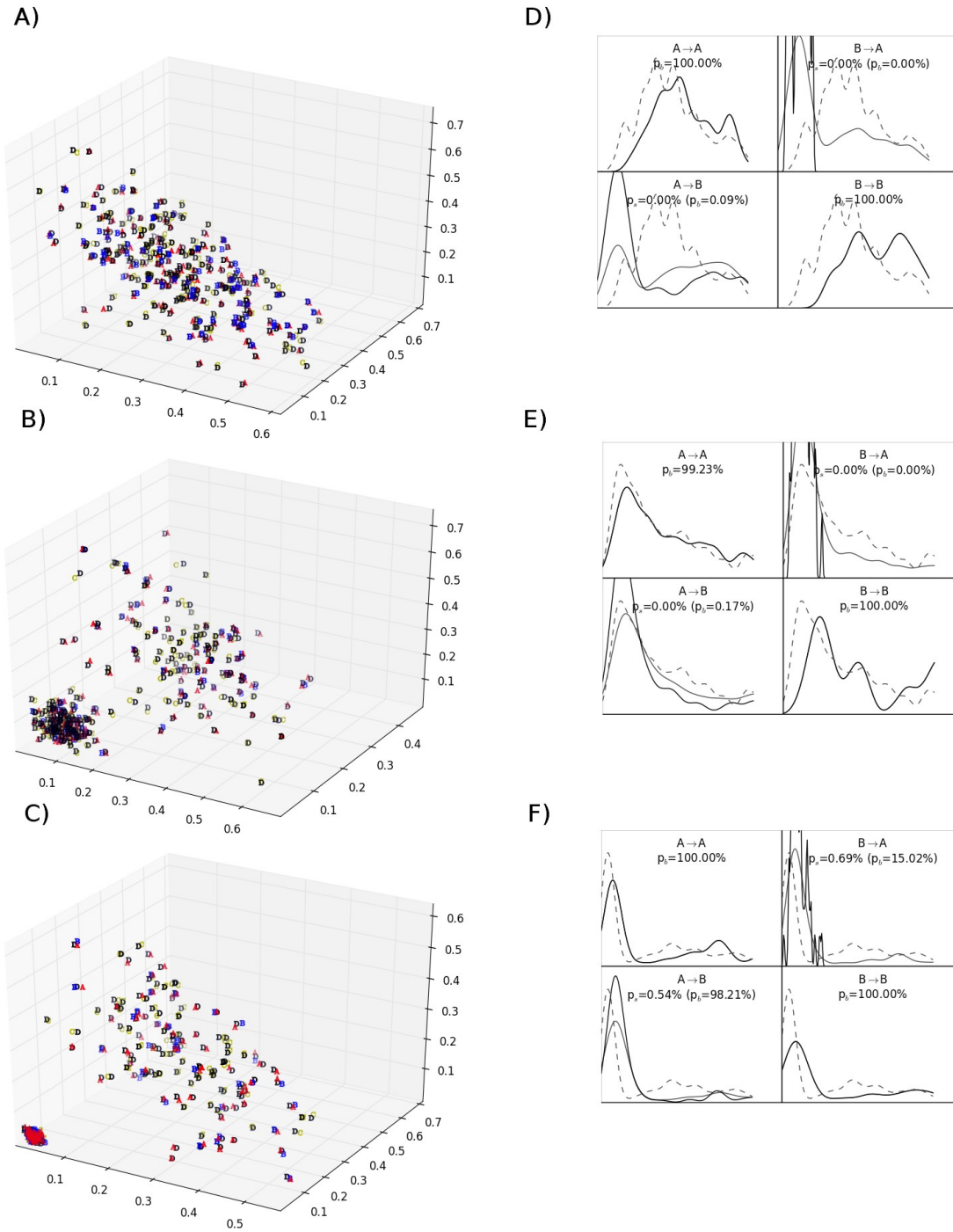


Figure 4.3: Randomly generated four-dimensional data under clustering scenarios of 1x, 4x and 16x (a-c) analyzed with the full protocol (d-f). Gene names correspond to gene names in Figures 4.1 and 4.2.



#### *4-D Example Visualizations*

Example four-dimensional data sets were prepared and analyzed by the full procedure to visually demonstrate how the method works. In all figures, the genes correspond to the genes in Figures 4.1 and 4.2). The locations of 256 hypothetical ORFs containing a gene D were chosen by generating four random coordinates from zero to one and normalizing them so their sum is one. Because the four dimensions sum to one, the fourth dimension is extraneous and the data can be displayed in a three-dimensional plot (Figure 4.3a-c). The process was then repeated, but half of the locations were compressed into, respectively, 1/64 and 1/4096 of their original volume (Figure 4.3b, 4.3c). Because some taxa will have more microdiversity than others and hence their contigs will occupy a tighter distribution of tetranucleotide space, real data is expected to be non-randomly distributed; this clustering example demonstrates how the method works under such conditions.

A (red letters) and C (yellow letters) are genes which are associated with half of all Ds, but are never found with each other. B (blue letters) occurs in half of all organisms that have a gene A (a quarter of all D's). The 4-dimensional locations of genes A, B, and C are determined by adding Gaussian noise with scale 0.005 to the location of their associated gene D.

In Figure 4.3d-f, nearest-neighbor distances from each gene to each other gene were computed as described above and their distribution plotted (black curve) as were the biological (dashed gray curve) and statistical (solid gray curve) controls. The one-sided Mann-Whitney U probabilities

that the gene-gene distances were greater than the distribution of the statistical control ( $p_s$ ) and biological control ( $p_b$ ) were calculated for all comparisons. The statistical control is not reported when the same gene is compared to itself (diagonal axes); the union of a gene and itself results in the same distribution as the unperturbed union and is hence identical to the original. The biological control need only be computed once per clustering scenario, so its curve (grey dashed) is the same in all plots for each scenario, despite differences in y-axis scaling.

## Results and Discussion

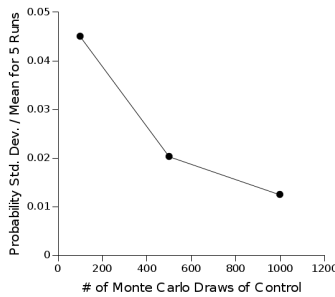
### *4-D Examples*

Although it can only be seen in the density plots under the most clustered scenario because both probabilities round to zero, the method yields closer association of  $B \rightarrow A$  than of  $A \rightarrow B$  (i.e.,  $p$  of  $B \rightarrow A$  is lower than  $p$  of  $A \rightarrow B$ ). By the standard metric of  $p < 0.05$ , both are also always significant and indicative of the expected association for the hypothetical scenario, i.e., all Bs are linked strongly to As, but not all As are linked strongly to Bs. As the clustering increases and half of all the genes are increasingly close to each other, the associations between A and B become weaker, although they do not disappear until all the points are unrealistically compressed to a space of around 1/2,000,000 their original volume (not shown).

Additionally, although A-C, B-C, and C-C comparisons were performed, as expected there was no significant association between C and the other genes. This non-association also holds true under the scenarios of increased clustering, demonstrating that while the method may produce

Type II statistical errors (false negatives), it is relatively immune to Type I statistical errors (false positives).

#### *Sensitivity Test of Model Parameters*



*Figure 4.4: Variance in probabilities of linkage as a result of changing number of Monte Carlo draws for the statistical control.*

First we examined the Mahoney Lake 7 meter metagenomic data under as many combinations of method parameters as possible (Table 4.2). The number of Monte Carlo replicates used to generate the statistical control was fixed at 500. Testing showed that this value was high enough to ensure the statistical control had asymptotically approached a final distribution and that it would on average result in no more than a 2% variance in the final probabilities (Figure 4.4).

*Table 4.2: Parameters tested in sensitivity analysis.*

Parameter	Tested Values	Description
Minimum HMMER E-value	$10^{-5}$ , $10^{-10}$ , $10^{-15}$	Lower values increase stringency of gene identifications, but also decrease total number of identified ORFs to be compared
Minimum # of Genes	5, 10, 25	Lowering increases number of ORFs compared, but may be vulnerable to false associations.
Minimum Contig Length	1000, 2000, 5000	Lowering increases number of genes to compare, but causes higher variability (lower accuracy) of tetranucleotide scores.

The first parameter tested was the E-value cutoff used by HMMER to assign gene identities to ORFs. Although  $10^{-10}$  is believed to be a robust choice (Kelley et al., 2003), lowering the cutoff

will both increase the number of “true” genes found, but will also increase the number of weakly homologous, but functionally unrelated genes and thus the number of false positives. The second parameter was the minimum contig length; all shorter sequences were eliminated from the metagenome before proceeding. Although 2000 base pairs is believed to generate representative tetranucleotide percentages, we also examined cutoffs of 1000 and 5000. A lower value allows many more contigs to be retained and thus more instances of each gene to be compared, but it decreases the accuracy of the assigned tetranucleotide profile and may cause spurious results. The third parameter was the minimum required number of ORFs identified a given gene; a number smaller than this cutoff eliminates the gene from the comparison list. Although a lower number once again allows more gene-gene associations to be probed, it also increases the chance of a false positive. For example, if only one instance of a given gene were present, it would have a high probability of randomly being found with any other gene and thus generating a spurious linkage.

*Table 4.3: Total number of identified genes from the Mahoney Lake 7M data set compared under different sets of parameters.*

<b>E-value</b>	<b>10e-5</b>	<b>10e-5</b>	<b>10e-5</b>	<b>10e-10</b>	<b>10e-10</b>	<b>10e-10</b>	<b>10e-15</b>	<b>10e-15</b>	<b>10e-15</b>
<b>Min. Contig Length</b>	<b>1000</b>	<b>2000</b>	<b>5000</b>	<b>1000</b>	<b>2000</b>	<b>5000</b>	<b>1000</b>	<b>2000</b>	<b>5000</b>
<b>Total Genes</b>	16432	13186	9337	9690	7900	5651	6244	5136	3703
<b>&gt;5 Copies</b>	16420	13174	9320	9681	7888	5636	6235	5120	3692
<b>&gt;10 Copies</b>	16398	13139	9299	9627	7838	5557	6192	5084	3618
<b>&gt;25 Copies</b>	16204	12958	9153	9496	7695	5461	5961	4876	3436

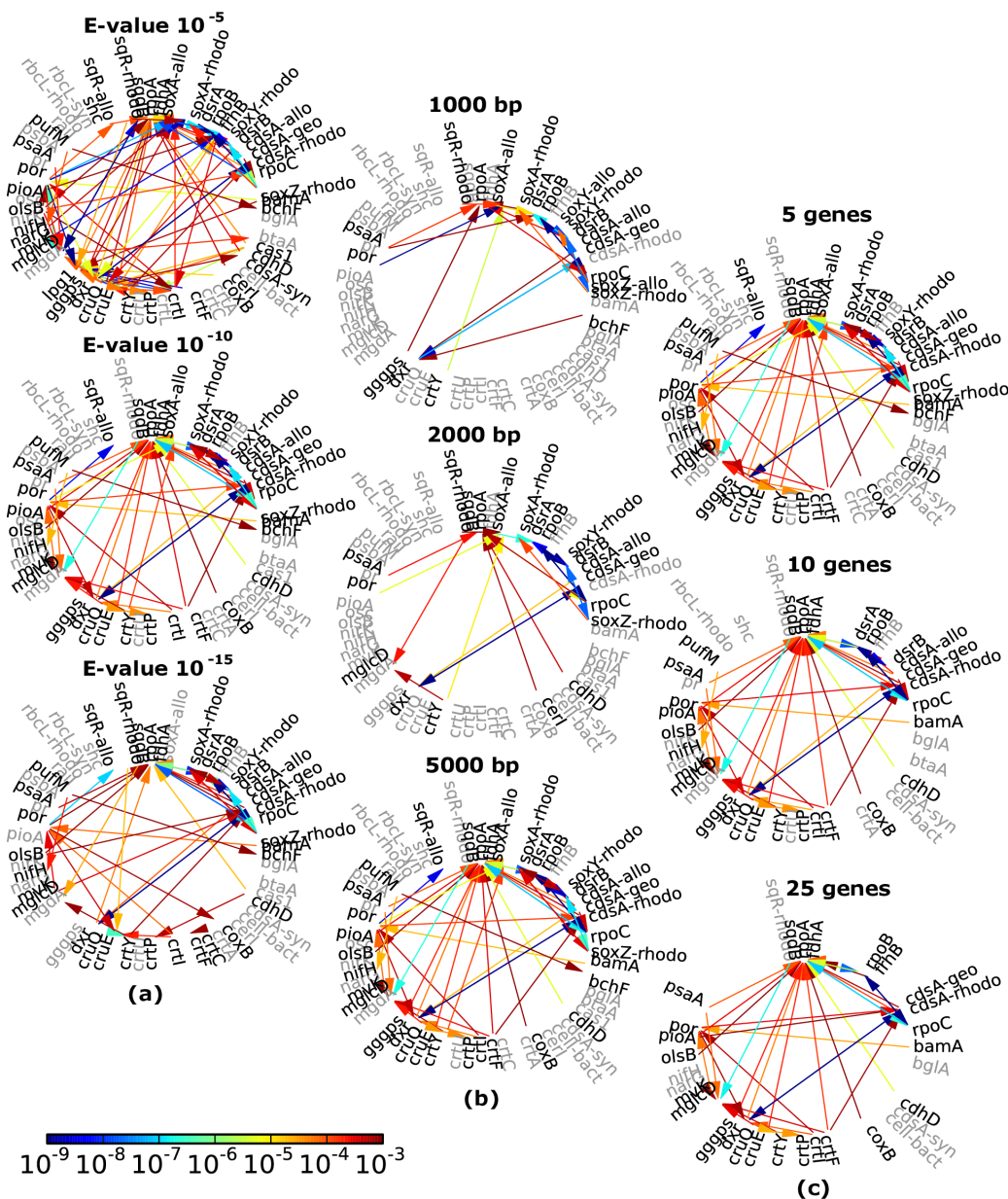


Figure 4.5: Associations between genes under varying parameters in the 7-m sample. Unless labeled otherwise, all associations are between at least 5 genes identified with e-values of  $10^{-10}$  on contigs with at least 5000 basepairs. Arrows indicate direction of association:  $A \rightarrow B$  implies if  $A$  is present,  $B$  will frequently be present also. (a) Varying e-value used for gene identification, (b) varying minimum contig size, and (c) varying minimum number of genes for identification. Gene names in black were found to have associations with other genes; names in gray have sufficient copy numbers for calculation under the criteria for that trial, but no associations; “blank” gene names did not appear frequently enough to be calculated.

### *Sensitivity Test Results*

Because Mahoney Lake is a low complexity ecosystem with few major species (Overmann et al., 1991; Hamilton et al., in prep.), the default choice was to use five genes as a minimum number for comparison. The minimum number of genes used (5, 10, 25) to calculate a linkage has little significant impact on the number or strength of linkages found. Increasing the number of required occurrences of each gene does not change existing linkages (Figure 4.5c), but it does remove several genes, such as sulfur oxidation (*sox*) and bacteriochlorophyll synthesis (*bchF*) genes that are likely important in lake processes. Based on this result, we retain 5 as the default choice for this parameter.

*Table 4.4: Number of contigs greater than a specific length and number of ORFs found on those contigs*

	<b>Contigs</b>	<b>ORFs</b>
<b>&gt;1000 bp</b>	125731	541611
<b>&gt;2000 bp</b>	59201	401649
<b>&gt;5000 bp</b>	19914	256724

Decreasing the E-value threshold to  $10^{-5}$  revealed some interesting links, such as from archaeal lipid synthesis (*gggps*) to sulfolipid synthesis (*sqdB*), an association which also has recently been reported in the literature (Meyer et al., 2011; Villanueva et al., 2013). However, at  $E = 10^{-5}$ , apparent false positives (homologs possibly of different function) also were detected. For example, fungal lignin peroxidase (*lpgI*; not believed to be a central or widely-distributed

process in this anoxic lake) gained widespread linkage from other genes, possibly because a homologous, but functionally different gene was identified as *lpgI*. As the number of genes identified was roughly similar between the  $10^{-10}$  and  $10^{-15}$  tests but markedly different from the  $10^{-5}$  test (Table 4.3), for further runs we decided to use an E-value of  $10^{-10}$ . Results from tests on different E-values are presented in an abbreviated format in Figure 4.5a.

There are less than one sixth the number of >5000 bp contigs as there are >1000 bp contigs (Table 4.4), but only slightly less than one half the number of ORFs were retrieved when examining these longer contigs (Table 4.3). In addition, a larger number of significant linkages was found when using this smaller number of longer contigs ( $\geq 5000$  basepairs; Figure 4.5b). This effect likely arises because longer contigs will yield tetranucleotide scores that more accurately represent the underlying organisms, thus improving the ability of the algorithm to detect linkages. For all further work, we examined >5000 bp contigs to ensure most accurate tetranucleotide scores for comparison.

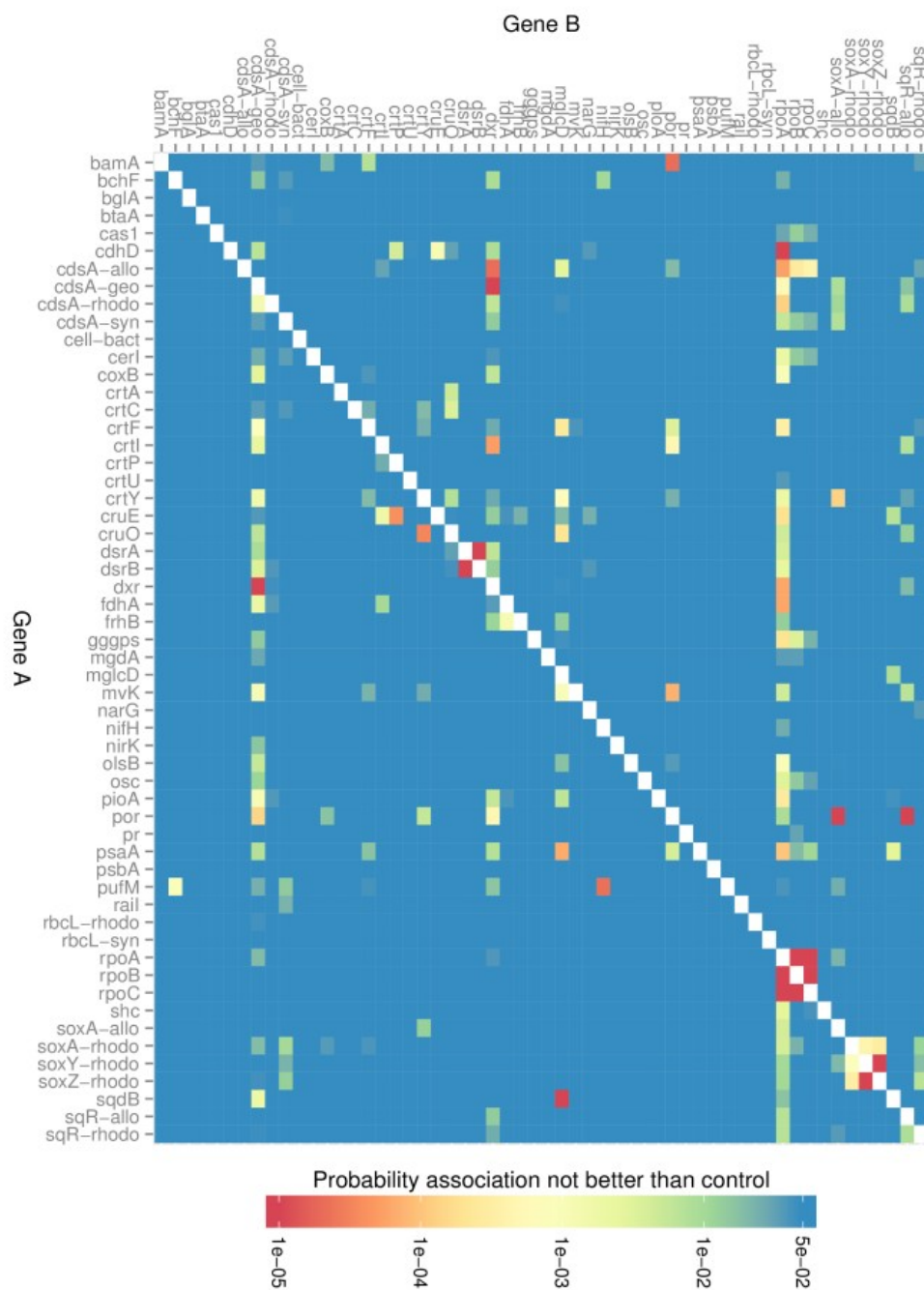


Figure 4.6: Heatmap showing associations between genes at 7-m with at least 5 copies, identified at E-value of  $10^{-10}$ , in contigs of length 5000 bp or greater. Linkages are directional from Gene A  $\rightarrow$  Gene B.



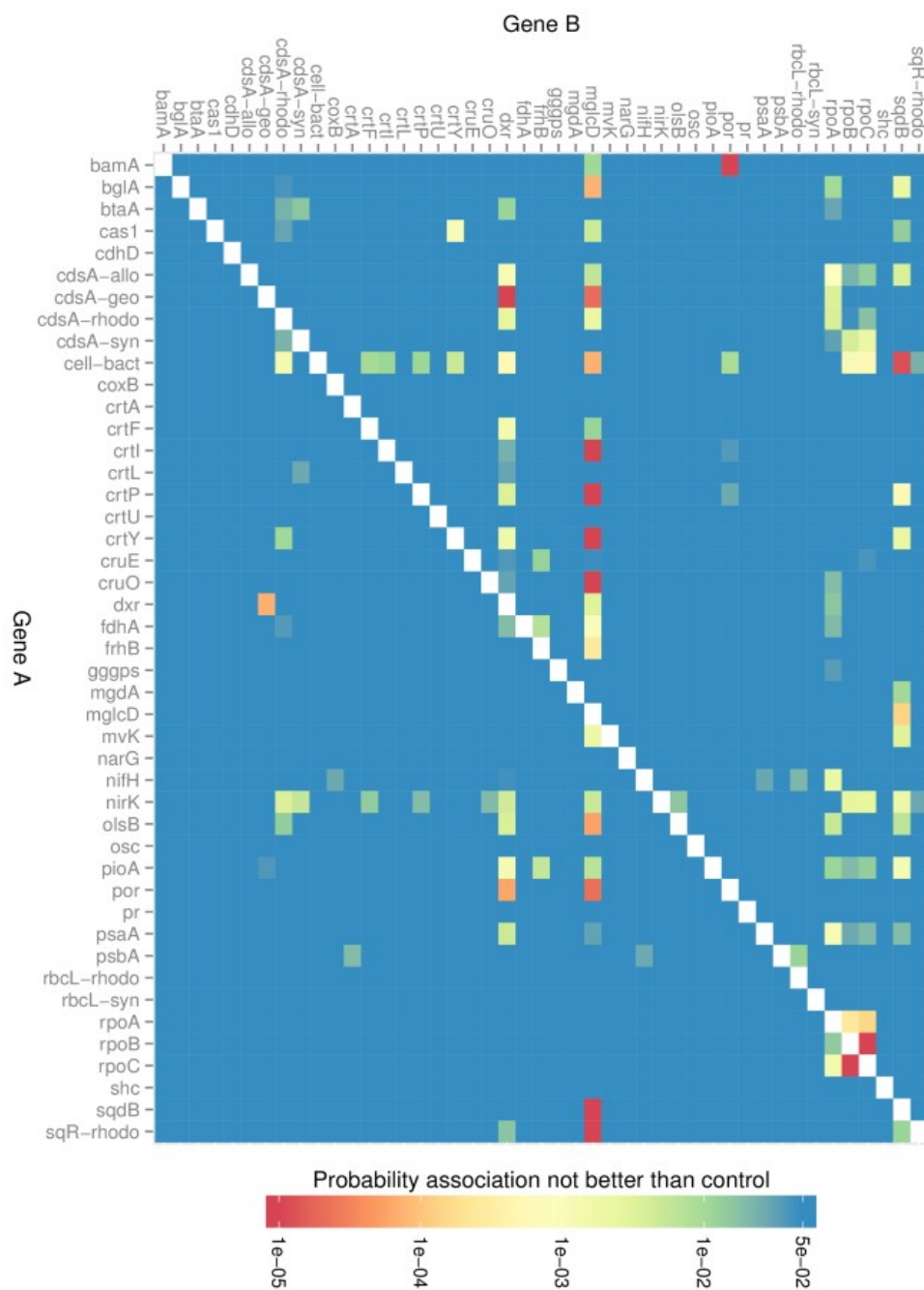


Figure 4.7: Heatmap showing associations between genes at 8-m with at least 5 copies, identified at e-value of  $10^{-10}$ , in contigs of length 5000 bp or greater. Linkages are directional from Gene A  $\rightarrow$  Gene B.

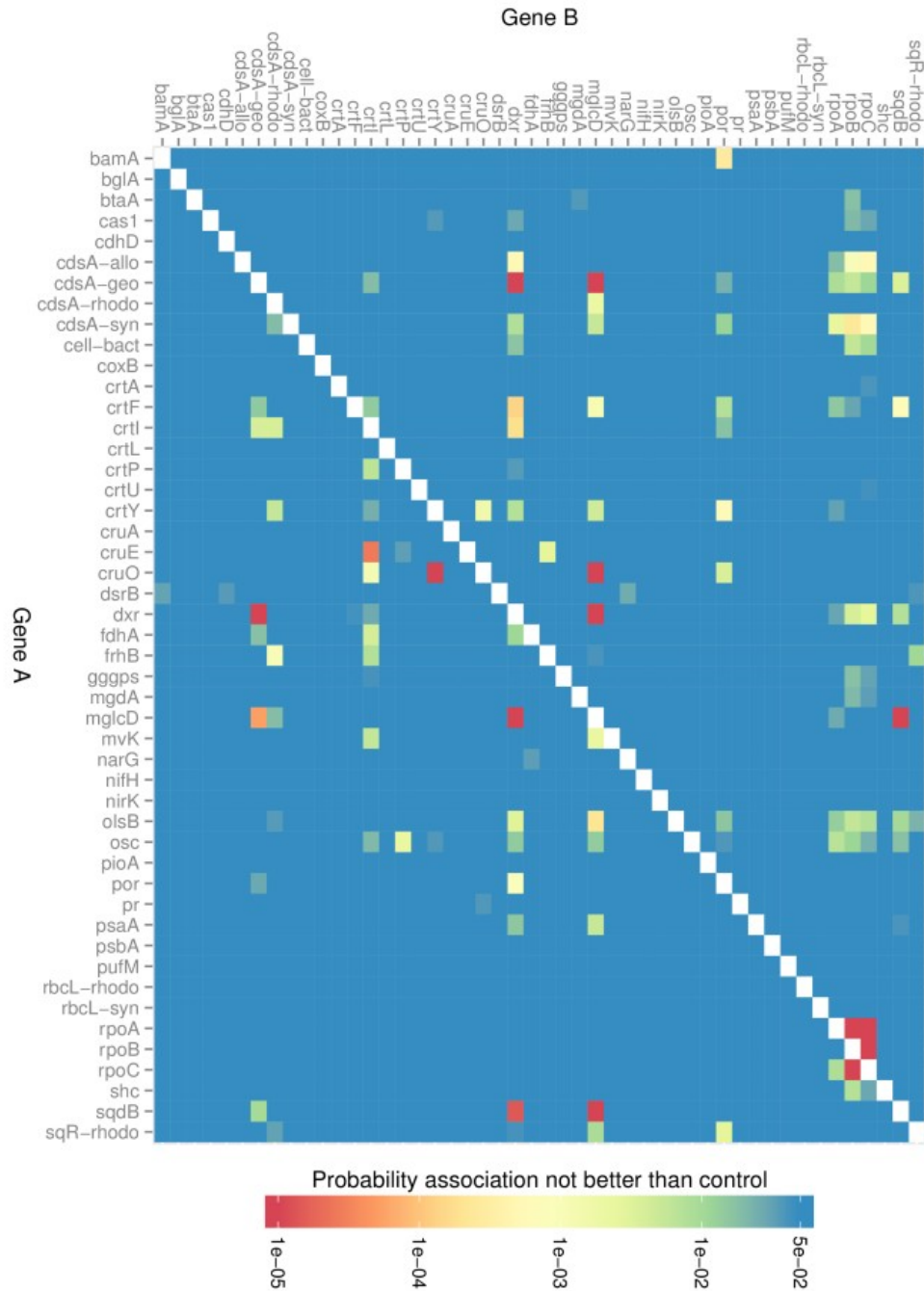


Figure 4.8: Heatmap showing associations between genes in lake bottom sediment with at least 5 copies, identified at e-value of  $10^{-10}$ , in contigs of length 5000 bp or greater. Linkages are directional from Gene A  $\rightarrow$  Gene B.

### *Results for Mahoney Lake Metagenomes*

Using the best parameters as determined above, results were calculated for the full gene list (Table 4.1) for the three Mahoney Lake metagenomes (Figures 4.6–4.8). Encouragingly, several expected gene-gene linkages were detected, while known false associations did not appear. Ribosomal protein genes (*rpoA*, *rpoB*, and *rpoC*) were linked to each other in all datasets, although *rpoA* was more weakly linked to *rpoB* and *rpoC* than they were to each other. In many microbes, *rpoB* and *rpoC* form an operon, while *rpoA* is located distally on the chromosome; e.g., in *Escherichia coli* str. *K-12 substr. MG1655* this operon is located 0.74 Mbp from the *rpoA* gene, increasing the possibility that it lies in a region with a slightly different tetranucleotide score. In a metagenomic dataset, the odds of having a 5000 bp contig containing both *rpoB* and *rpoC* are much greater than having a 0.74 Mbp contig with all three genes. Thus the expected probability of linkage from *rpoA* to the *rpoBC* operon is lower, although it still should (and does) detect a significant linkage due to the inherent similarity in tetranucleotide scores.

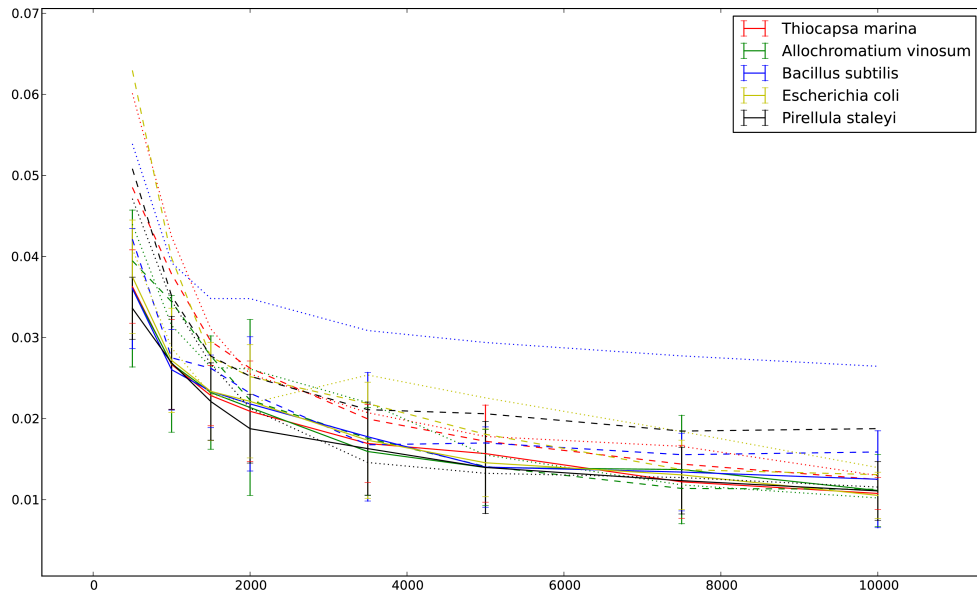


Figure 4.9: Relationship between tetranucleotide scores of *rpo* genes and whole genomes for five genomes. The Y-axis is the Euclidean distance between (1) the tetranucleotide score of a section of DNA centered on either the middle of a *rpoA* (dashed) or *rpoB* (dotted) gene or on another random place in the genome (solid line with error bars representing standard deviation of 50 samples) and (2) the tetranucleotide score of the whole genome. The X-axis is the number of nucleotides in that given section.

Additionally, all functional genes should link significantly to *rpoA*, *rpoB*, and *rpoC*; yet the results show that many functional genes do not have low probability scores (significant positive linkages) when compared to these genes, especially for *rpoBC* (Figures 4.6-4.8). There are two simple, possible explanations for these lower than expected linkage strengths. (1) Ribosomal proteins are essential for cellular processes and are thus highly conserved; this conservation may manifest itself as a tetranucleotide signature distinct from the rest of the genome. One way to test this hypothesis is to examine if the Euclidean distance from the tetranucleotide score of a hypothetical contig containing an *rpo* gene to the tetranucleotide score of the whole genome is

greater than the distance from any other sequence of the same length to the whole genome. Tests on five model genomes reveal that indeed *rpo* genes are more distant in tetranucleotide space, although not by more than two or three standard deviations of normal genomic heterogeneity (Figure 4.9). (2) Alternatively, as our metagenomes lack full sequence coverage, it is also possible that many of the functional genes either lack a corresponding ribosomal protein in the overall dataset, and/or the ribosomal proteins did not assemble well and were filtered out by the contig length cutoff discussed above. In the 7-m dataset there were 126 *rpoA* genes and 169 *rpoB* genes identified at the  $10^{-10}$  threshold in contigs longer than 2000 basepairs, and just 94 and 114, respectively, in contigs > 5000 bp. These numbers are much lower than a previous estimate of 222-934 OTUs derived from previous work (Klepac-Ceraj et al., 2012) indicating an incomplete sampling of the microbial diversity of this system.

Some genes were not found in enough copies ( $\geq 5$ ) to make comparisons in any sample, including iron (*iro*) and manganese (*ompC*) oxidation. Iron and manganese levels in the lake should be low due to the high concentrations of sulfide, so this is also to be expected. Similarly, not enough instances of bacterial or archaeal nitrification pathways (*amoA*) were found.

Sulfate reduction genes (*dsrA* and *dsrB*) are only found in the 7-m layer and are bidirectionally linked, as expected. Similarly, sulfur oxidation genes (*soxA*, *soxY*, and *soxZ*) are also only found as a group in the 7-m dataset, and they also form a completely bidirectionally linked group. Unfortunately, other sets of known gene pairs were not found at  $\geq 5$  copies (*pufM* and *pufL*, *psaA* and *psaB*, *psbA* and *psbB*), likely because of the relatively low diversity of photosynthetic

species in this system (Hamilton et al., 2013). The lack of association between the few oxygenic photosystem I and II genes which were found (*psaA* and *psbA*) is likely due to incomplete metagenome coverage, especially as DNA from oxygenic photoautotrophs would be expected to come from “fossil” sinking organic particles. An indication of the general absence of false positives comes from genes that are functionally redundant (e.g., *puf* versus *psa* photosystem genes, or *dxr* versus *mvk* isoprenoid biosynthetic pathways): these pairs were not found to be linked in any sample.

There are, however, a few associations that may reflect false positives, and this may be due to the incorrect assignment of gene identity to a homologous sequence encoding for a different function. For example, in the 7-m data, anaerobic degradation of aromatic compounds (*bamA*) is linked to a light-dependent chlorophyll synthesis gene specific to oxygenic phototrophs (*por*). This chlorophyll synthesis gene is, in turn, linked to many genes including both *soxA* (Allochromatium and Rhodopseudomonas homologues) and phototrophic sulfide oxidation (*sqR*). Oxygenic photoautotrophs containing chlorophyll are unlikely to be anaerobically degrading aromatic compounds in these samples. Also, although chlorophyll-containing cyanobacteria can oxidize sulfide (Cohen et al., 1986; Johnston et al., 2009), both of these results are more likely to have been caused by false inclusion of short-chain dehydrogenase genes homologous to *por* that are found in bacteria such as *Mycobacterium* and *Streptomyces* spp..

A few genes had associations with many other genes. These were primarily phospholipid synthesis (*cdsA*), “cyanobacterial” glycolipid synthesis (*mgldD*), and the non-mevalonate or

deoxyxylulose 5-phosphate isoprenoid synthesis pathway (*dxr*). All are expected to be important for lipid biosynthesis and are widespread in bacteria. This fits with the predominance of Chromatiaceae throughout the lake (Overmann et al., 1991; Overmann et al., 1992) and Gammaproteobacteria in general making up the majority of genus-level diversity at 7 m (Klepac-Ceraj et al., 2012). Chromatiaceae use the non-mevalonate pathway (Eisenreich et al., 2004) and produce both phospholipids and glycolipids (Imhoff and Bias-Imhoff, 1995). By comparison, the glycolipid synthesis gene from *Chlorobi* (*mgdA*) was generally not associated with other genes.

The genes *dxr* and *cdsA* themselves were bidirectionally closely associated, consistent with their common occurrence in a single transcriptional unit (Grolle et al., 2000; Takahashi et al., 1998). In a study of the intact polar lipids of a marine water column, Van Mooy and Fredricks (2010) found that phospholipids were dominantly associated with heterotrophic bacteria but were also present in small quantities in photosynthetic organisms containing thylkaloid membranes, both of which are likely to produce isoprenoids via the *dxr* pathway (Boucher and Doolittle, 2000). In addition to the *dxr* isoprenoid pathway, the gene for the mevalonate kinase pathway (*mvk*) was also found in all of the samples. Its strongest association was to the chlorophyll synthesis gene (*por*) at 7 m, possibly reflecting use of the mevalonate kinase pathway in plants (Rohmer, 1999). The overwhelming dominance of the non-mevalonate pathway in Mahoney Lake does not necessarily reflect all microbial ecosystems, and other environments should be studied to examine how the distribution of isoprenoid pathways changes between environments.

The glycolipid synthesis gene from cyanobacteria (*mglcD*) was bidirectionally linked to sulfolipid synthesis (*sqdB*). Sequenced purple sulfur bacterial genomes are not known to contain sulfolipid synthesis genes (Villanueva et al., 2013) and no definitive identifications of sulfolipids have been made in the Chromatiaceae (Imhoff and Bias-Imhoff, 1995), so these associations are believed to be from Cyanobacteria or other oxygenic photoautotrophs sinking from the mixolimnion (Klepac-Ceraj et al., 2012). As most cyanobacterial sulfolipids contain both sulfonyl and glucose moieties (Riekhof et al., 2003, Van Mooy et al., 2006), *sqdB* may require *mglcD*, but not necessarily the reverse. In agreement, the relative strength of association from *sqdB* to *mglcD* is much stronger than the reverse in our results.

Consistent with the hypothesis that sulfolipid genes are predominantly from oxygenic photoautotrophs, there is a link from a photosystem I gene (*psaA*) to *sqdb*, and *psaA* is also linked to *mglcD*. Associations to *cdsA* were also strongest in the 7 m layer, while associations to *mglcD* were strongest in the 8 m layer, potentially indicating that a greater proportion of DNA in this layer derives from sinking remains of oxygenic photoautotrophs which more commonly produce glycolipids (Hölzl & Dörmann, 2007). Steroid synthesis genes (*osc*; *casI*), expected to be primarily from eukaryotes, had no associations at 7 m, but in the sediment were found associated with the carotenoid synthesis genes required for beta-carotene (*crtI*, *crtP*, *crtY*) and to *sqdb*. Hopanoid synthesis genes (*shc*) were present in all samples, but no associations were found between them and any other genes (except *rpo*). This is consistent with suggestions that hopanoid synthesis is neither phylogenetically nor metabolically associated to any specific bacteria (Pearson et al., 2007).



The one gene associated with bacteriochlorophyll synthesis (*bchF*) is the proteobacterial photosystem gene (*pufM*); this is consistent with the production of bacteriochlorophyll in this purple sulfur bacteria-dominated ecosystem. *pufM* also has an association to nitrogen fixation, which suggests that the Chromatiaceae in Mahoney Lake are able to fix nitrogen (Asao et al., 2007). However, the reverse association is not observed to be true; *nifH* is linked to the common lipid biosynthesis genes (*cdsA*, *dxr*) and *soxA-allo*, potentially indicating that nitrogen fixation in this environment is more generally performed by organisms involved in sulfur oxidation.

Bacterial cellulase (*cel*, labelled “cell-bact”) is linked to a number of genes at 8 m, as is the gene for nitrite reduction (*nirK*). These linkages may be representative of a single species or closely related taxa with a stable set of genes performing these processes in the anoxic monimolimnion.

No associations between carotenoid genes were found at 8 m, but interesting associations existed at both 7 m and in the sediment. At 7 m, a methylase (*crtF*) was bidirectionally linked to lycopene cyclase (*crtY*) which in turn was bidirectionally linked to a carotenoid ketolase (*cruO*). A carotenoid hydroxylase (*crtC*) was in turn directionally linked to all of these genes, but none were linked to it. All four of these genes are in the okenone-producing pathway of purple-sulfur bacteria (Vogl and Bryant, 2011; 2012) and may be expected to co-occur in the lake. A carotenoid monooxygenase (*crtA*) was also directionally linked to *cruO*. *crtA* is known to ketolase spheroidene and spheroidenone, but no close homologues have previously been found in okenone-producing bacteria (Vogl and Bryant, 2011). Another ketolase (*cruS*) is present in the

species investigated by Vogl and Bryant (2011); it is possible that *crtA* is replacing its function in this environment.

The bidirectional linkage between *cruO* and *crtY* was also present in the sediment. Many more genes were associated to *crtI* in the sediment, indicative of its central role in initiating almost all carotenoid synthesis pathways. There was also a link from beta-carotene desaturase (*cruE*) to phytoene desaturase (*crtP*) at both 7 m and in the sediment. Both genes are found in cyanobacteria (*crtP* is sometimes annotated as *crtQ*; Graham, 2008) and are involved in the formation of renierapurpurin (Graham and Bryant, 2008), although this carotenoid was below detection in all samples of the lake (Chapter 2, this thesis).

#### *Final comments*

There are still some limitations with this method. It will not resolve situations where there are links between pathways in syntrophic organisms, nor will it resolve instances where the metabolisms of two organisms with different tetranucleotide frequencies are coupled. For example, the anaerobic oxidation of methane by sulfate seems to be mediated by a consortium of a bacterium and an archaeon. Despite their close proximity, these two microbes probably do not share the same tetranucleotide frequency, and thus associations between methane oxidation and sulfate reduction would not be found using these approaches.

Another limitation is this method only analyzes the genetic capacity of an environmental community. The presence of a gene does not necessarily imply its expression (Frias-Lopez et al., 2008). However, because tetranucleotide frequencies could be calculated on RNA in addition to

DNA, our method could be extended to the analysis of longer transcriptional fragments from metatranscriptomes. Although amino acid usage patterns in proteins are in part influenced by biases in nucleotide usage (Singer and Hickey, 2000), extending the method to peptide sequences is probably not feasible.

## Conclusion

We constructed a method to find associations between genes (e.g., their simultaneous presence in an organism or in strains bearing closely related patterns of tetranucleotide usage) using Euclidian distances between tetranucleotide frequencies. When applying this method to metagenomic samples from a meromictic lake, we revealed both previously known associations (e.g., *dxx* ↔ *cdsA*) and some novel associations (e.g., *crtA* → *cruO*). There is future potential to use this method on metatranscriptomes and on next-generation metagenomes from other sample sites. Using expanded, curated gene libraries for calculating the gene associations promises to refine and expand our knowledge of gene interactions and patterns in the environment.

## References

- Abe T, Kanaya S, Kinouchi M, Ichiba Y, Kozuki T, Ikemura T (2003) Informatics for Unveiling Hidden Genome Signatures. *Genome Research*, **13**(4), 693–702.
- Altschul SF, Madden TL, Schäffer AA, Zhang J, Zhang Z, Miller W, Lipman DJ (1997) Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Research*, **25**(17), 3389–3402.
- Andersen M (1992) Spatial analysis of two-species interactions. *Oecologia*, **91**(1), 134–140.

Asao M, Takaichi S, Madigan MT (2007) *Thiocapsa imhoffii*, sp. nov., an alkaliphilic purple sulfur bacterium of the family Chromatiaceae from Soap Lake, Washington (USA). *Archives of Microbiology*, **188**(6), 665–675.

Bhadra B, Raghukumar C, Pindi PK, Shivaji S (2008) *Brevibacterium oceani* sp. nov., isolated from deep-sea sediment of the Chagos Trench, Indian Ocean. *International Journal of Systematic and Evolutionary Microbiology*, **58**(1), 57–60.

Bohnebeck U, Lombardot T, Kottmann R, Glöckner FO (2008) MetaMine – A tool to detect and analyse gene patterns in their environmental context. *BMC Bioinformatics*, **9**(1), 459.

Boucher Y, Doolittle WF (2000) The role of lateral gene transfer in the evolution of isoprenoid biosynthesis pathways. *Molecular Microbiology*, **37**(4), 703–716.

Brocks JJ, Pearson A (2005) Building the biomarker tree of life. *Reviews in Mineralogy and Geochemistry*, **59**(1), 233.

Burke C, Steinberg P, Rusch D, Kjelleberg S, Thomas T (2011) Bacterial community assembly based on functional genes rather than species. *Proceedings of the National Academy of Sciences*, **108**(34), 14288–14293.

Cohen Y, Jørgensen BB, Revsbech NP, Poplawski R (1986) Adaptation to Hydrogen Sulfide of Oxygenic and Anoxygenic Photosynthesis among Cyanobacteria. *Applied and Environmental Microbiology*, **51**(2), 398–407.

Dick GJ, Andersson AF, Baker BJ, Simmons SL, Thomas BC, Yelton AP, Banfield JF (2009) Community-wide analysis of microbial genome sequence signatures. *Genome Biology*, **10**(8), R85.

Dinsdale EA, Edwards RA, Hall D, Angly F, Breitbart M, Brulc JM, Furlan M, Desnues C, Haynes M, Li L, McDaniel L, Moran MA, Nelson KE, Nilsson C, Olson R, Paul J, Brito BR, Ruan Y, Swan BK, Stevens R, Valentine DL, Thurber RV, Wegley L, White BA, Rohwer F (2008) Functional metagenomic profiling of nine biomes. *Nature*, **452**(7187), 629–632.

Eddy SR (2011) Accelerated Profile HMM Searches. *PLoS Comput Biol*, **7**(10), e1002195.

Edgar RC (2004) MUSCLE: multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Research*, **32**(5), 1792–1797.

Eisenreich W, Bacher A, Arigoni D, Rohdich F (2004) Biosynthesis of isoprenoids via the non-mevalonate pathway. *Cellular and Molecular Life Sciences CMLS*, **61**(12), 1401–1426.

Ferrer M, Golyshina OV, Chernikova TN, Khachane AN, Reyes-Duarte D, Santos VAPMD, Strompl C, Elborough K, Jarvis G, Neef A, Yakimov MM, Timmis KN, Golyshin PN (2005) Novel hydrolase diversity retrieved from a metagenome library of bovine rumen microflora. *Environmental Microbiology*, **7**(12), 1996–2010.

Frias-Lopez J, Shi Y, Tyson GW, Coleman ML, Schuster SC, Chisholm SW, DeLong EF (2008) Microbial community gene expression in ocean surface waters. *Proceedings of the National Academy of Sciences*, **105**(10), 3805–3810.

Frost LS, Leplae R, Summers AO, Toussaint A (2005) Mobile genetic elements: the agents of open source evolution. *Nature Reviews Microbiology*, **3**(9), 722–732.

Graham Joel E., Bryant DA (2008) The Biosynthetic Pathway for Synechoxanthin, an Aromatic Carotenoid Synthesized by the Euryhaline, Unicellular Cyanobacterium *Synechococcus* sp. Strain PCC 7002. *Journal of Bacteriology*, **190**(24), 7966–7974.

Graham Joel Edward (2008) *Carotenoid Biosynthesis in Synechococcus Sp. PCC 7002: Identification of the Enzymes and the Carotenoids*. ProQuest.

Gray ND, Head IM (2001) Linking genetic identity and function in communities of uncultured bacteria. *Environmental Microbiology*, **3**(8), 481–492.

Grice K, Cao C, Love GD, Twitchett RJ, Bottcher M, Grosjean E, Summons RE, Turgeon SC, Dunning W, Jin Y (2005) Photic Zone Euxinia During the Permian-Triassic Superanoxic Event. *Science*, **307**(5710), 706–709.

Grolle S, Bringer-Meyer S, Sahm H (2000) Isolation of the dxr gene of *Zymomonas mobilis* and characterization of the 1-deoxy-D-xylulose 5-phosphate reductoisomerase. *FEMS Microbiology Letters*, **191**(1), 131–137.

Hamilton TL, Bovee RJ, Sattin SR, Mohr W, Schaperdoth I, Gilhooly WP, Lyons TW, Pearson A, Macalady JL (2013) Sulfur dynamo in the phototrophic plate of a meromictic lake. *Geobiology*, (submitted).

Hess M, Sczyrba A, Egan R, Kim T-W, Chokhawala H, Schroth G, Luo S, Clark DS, Chen F, Zhang T, Mackie RI, Pennacchio LA, Tringe SG, Visel A, Woyke T, Wang Z, Rubin EM (2011) Metagenomic Discovery of Biomass-Degrading Genes and Genomes from Cow Rumen. *Science*, **331**(6016), 463–467.

Hölzl G, Dörmann P (2007) Structure and function of glycolipids in plants and bacteria. *Progress in Lipid Research*, **46**(5), 225–243.

Imhoff J, Bias-Imhoff U (1995) Lipids, quinones and fatty acids of anoxygenic phototrophic bacteria. In: *Anoxygenic Photosynthetic Bacteria* (eds. Blankenship R, Madigan M, Bauer C). Springer, Dordrecht, pp. 179–205.

Ishoei T, Woyke T, Stepanauskas R, Novotny M, Lasken RS (2008) Genomic sequencing of single microbial cells from environmental samples. *Current Opinion in Microbiology*, **11**(3), 198–204.

Iverson V, Morris RM, Frazar CD, Berthiaume CT, Morales RL, Armbrust EV (2012) Untangling Genomes from Metagenomes: Revealing an Uncultured Class of Marine Euryarchaeota. *Science*, **335**(6068), 587–590.

Jiang B, Song K, Ren J, Deng M, Sun F, Zhang X (2012) Comparison of metagenomic samples using sequence signatures. *BMC Genomics*, **13**(1), 730.

Johnston DT, Wolfe-Simon F, Pearson A, Knoll AH (2009) Anoxygenic photosynthesis modulated Proterozoic oxygen and sustained Earth's middle age. *Proceedings of the National Academy of Sciences*, **106**(40), 16925–16929.

Jones DS, Albrecht HL, Dawson KS, Schaperdorth I, Freeman KH, Pi Y, Pearson A, Macalady JL (2011) Community genomic analysis of an extremely acidophilic sulfur-oxidizing biofilm. *The ISME Journal*, **6**(1), 158–170.

Kalyuzhnaya MG, Lapidus A, Ivanova N, Copeland AC, McHardy AC, Szeto E, Salamov A, Grigoriev IV, Suciú D, Levine SR, Markowitz VM, Rigoutsos I, Tringe SG, Bruce DC, Richardson PM, Lidstrom ME, Chistoserdova L (2008) High-resolution metagenomics targets specific functional types in complex microbial communities. *Nature Biotechnology*, **26**(9), 1029–1034.

Karlin S, Ladunga I, Blaisdell BE (1994) Heterogeneity of genomes: measures and values. *Proceedings of the National Academy of Sciences*, **91**(26), 12837–12841.

Kelley BP, Sharan R, Karp RM, Sittler T, Root DE, Stockwell BR, Ideker T (2003) Conserved pathways within bacteria and yeast as revealed by global protein network alignment. *Proceedings of the National Academy of Sciences*, **100**(20), 11394–11399.

Klepac-Ceraj V, Hayes CA, Gilhooly WP, Lyons TW, Kolter R, Pearson A (2012) Microbial diversity under extreme euxinia: Mahoney Lake, Canada. *Geobiology*, 223–235.

Kohl W, Achenbach H, Reichenbach H (1983) The pigments of *Brevibacterium linens*: Aromatic carotenoids. *Phytochemistry*, **22**(1), 207–210.

Lohse M, Bolger AM, Nagel A, Fernie AR, Lunn JE, Stitt M, Usadel B (2012) RobiNA: a user-friendly, integrated software solution for RNA-Seq-based transcriptomics. *Nucleic Acids Research*, **40**(W1), W622–W627.

Mavromatis K, Chu K, Ivanova N, Hooper SD, Markowitz VM, Kyrpides NC (2009) Gene Context Analysis in the Integrated Microbial Genomes (IMG) Data Management System. *PLoS ONE*, **4**(11), e7979.

McHardy AC, Rigoutsos I (2007) What's in the mix: phylogenetic classification of metagenome sequence samples. *Current Opinion in Microbiology*, **10**(5), 499–503.

Medini D, Donati C, Tettelin H, Massignani V, Rappuoli R (2005) The microbial pan-genome. *Current Opinion in Genetics & Development*, **15**(6), 589–594.

Meyer BH, Zolghadr B, Peyfoon E, Pabst M, Panico M, Morris HR, Haslam SM, Messner P, Schäffer C, Dell A, Albers S-V (2011) Sulfoquinovose synthase – an important enzyme in the N-glycosylation pathway of *Sulfolobus acidocaldarius*. *Molecular Microbiology*, **82**(5), 1150–1163.

Overbeek R, Fonstein M, D'Souza M, Pusch GD, Maltsev N (1999) The use of gene clusters to infer functional coupling. *Proceedings of the National Academy of Sciences*, **96**(6), 2896–2901.

Overmann J, Beatty JT, Hall KJ, Pfennig N, Northcote TG (1991) Characterization of a dense, purple sulfur bacterial layer in a meromictic salt lake. *Limnology and Oceanography*, **36**(3), 846–859.

Overmann J, Fischer U, Pfennig N (1992) A new purple sulfur bacterium from saline littoral sediments, *Thiorhodovibrio winogradskyi* gen. nov. and sp. nov. *Archives of Microbiology*, **157**(4), 329–335.

Pearson A, Page SRF, Jorgenson T, Fischer W, Higgins MB (2007) Novel hopanoid cyclases from the environment. *Environmental Microbiology*, 2175–2188.

Peng Y, Leung HCM, Yiu SM, Chin FYL (2012) IDBA-UD: a de novo assembler for single-cell and metagenomic sequencing data with highly uneven depth. *Bioinformatics*, **28**(11), 1420–1428.

Perry SC, Beiko RG (2010) Distinguishing Microbial Genome Fragments Based on Their Composition: Evolutionary and Comparative Genomic Perspectives. *Genome Biology and Evolution*, **2**(0), 117–131.

Pester M, Bittner N, Deevong P, Wagner M, Loy A (2010) A “rare biosphere” microorganism contributes to sulfate reduction in a peatland. *The ISME Journal*, **4**(12), 1591–1602.

Podar M, Abulencia CB, Walcher M, Hutchison D, Zengler K, Garcia JA, Holland T, Cotton D, Hauser L, Keller M (2007) Targeted Access to the Genomes of Low-Abundance Organisms in Complex Microbial Communities. *Applied and Environmental Microbiology*, **73**(10), 3205–3214.

Pride DT, Meinersmann RJ, Wassenaar TM, Blaser MJ (2003) Evolutionary Implications of Microbial Genome Tetranucleotide Frequency Biases. *Genome Research*, **13**(2), 145–158.

Quaiser A, Zivanovic Y, Moreira D, Lopez-Garcia P (2010) Comparative metagenomics of bathypelagic plankton and bottom sediment from the Sea of Marmara. *The ISME Journal*, **5**(2), 285–304.

R Development Core Team (2011) R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. <http://www.R-project.org/>.



Raes J, Bork P (2008) Molecular eco-systems biology: towards an understanding of community function. *Nature Reviews Microbiology*, **6**(9), 693–699.

Raghunathan A, Ferguson HR, Bornarth CJ, Song W, Driscoll M, Lasken RS (2005) Genomic DNA Amplification from a Single Bacterium. *Applied and Environmental Microbiology*, **71**(6), 3342–3347.

Riekhof WR, Ruckle ME, Lydic TA, Sears BB, Benning C (2003) The Sulfolipids 2'-O-Acyl-Sulfoquinovosyldiacylglycerol and Sulfoquinovosyldiacylglycerol Are Absent from a *Chlamydomonas reinhardtii* Mutant Deleted in SQD1. *Plant Physiology*, **133**(2), 864–874.

Rohmer M (1999) The discovery of a mevalonate-independent pathway for isoprenoid biosynthesis in bacteria, algae and higher plants. *Natural Product Reports*, **16**(5), 565–574.

Saeed I, Tang S-L, Halgamuge SK (2012) Unsupervised discovery of microbial population structure within metagenomes using nucleotide base composition. *Nucleic Acids Research*, **40**(5), e34.

Shade A, Handelsman J (2012) Beyond the Venn diagram: the hunt for a core microbiome. *Environmental Microbiology*, **14**(1), 4–12.

Sharon I, Morowitz MJ, Thomas BC, Costello EK, Relman DA, Banfield JF (2012) Time series community genomics analysis reveals rapid shifts in bacterial species, strains, and phage during infant gut colonization. *Genome Research*,

Singer GAC, Hickey DA (2000) Nucleotide Bias Causes a Genomewide Bias in the Amino Acid Composition of Proteins. *Molecular Biology and Evolution*, **17**(11), 1581–1588.

Sogin ML, Morrison HG, Huber JA, Welch DM, Huse SM, Neal PR, Arrieta JM, Herndl GJ (2006) Microbial diversity in the deep sea and the underexplored “rare biosphere.” *Proceedings of the National Academy of Sciences*, **103**(32), 12115–12120.

Suenaga H (2012) Targeted metagenomics: a high-resolution metagenomics approach for specific gene clusters in complex microbial communities. *Environmental Microbiology*, **14**(1), 13–22.

Suwanto A, Kaplan S (1989) Physical and genetic mapping of the *Rhodobacter sphaeroides* 2.4.1 genome: presence of two unique circular chromosomes. *Journal of Bacteriology*, **171**(11), 5850–5859.

Swingley WD, Meyer-Dombard DR, Shock EL, Alsop EB, Falenski HD, Havig JR, Raymond J (2012) Coordinating Environmental Genomics and Geochemistry Reveals Metabolic Transitions in a Hot Spring Ecosystem. *PLoS ONE*, **7**(6), e38108.

Takahashi S, Kuzuyama T, Watanabe H, Seto H (1998) A 1-deoxy-d-xylulose 5-phosphate reductoisomerase catalyzing the formation of 2-C-methyl-d-erythritol 4-phosphate in



an alternative nonmevalonate pathway for terpenoid biosynthesis. *Proceedings of the National Academy of Sciences*, **95**(17), 9879–9884.

Teeling H, Waldmann J, Lombardot T, Bauer M, Glöckner FO (2004a) TETRA: a web-service and a stand-alone program for the analysis and comparison of tetranucleotide usage patterns in DNA sequences. *BMC Bioinformatics*, **5**(1), 163.

Teeling H, Meyerdierks A, Bauer M, Amann R, Glöckner FO (2004b) Application of tetranucleotide frequencies for the assignment of genomic fragments. *Environmental Microbiology*, **6**(9), 938–947.

Tringe SG, Mering C von, Kobayashi A, Salamov AA, Chen K, Chang HW, Podar M, Short JM, Mathur EJ, Detter JC, Bork P, Hugenholtz P, Rubin EM (2005) Comparative Metagenomics of Microbial Communities. *Science*, **308**(5721), 554–557.

Tyson GW, Chapman J, Hugenholtz P, Allen EE, Ram RJ, Richardson PM, Solovyev VV, Rubin EM, Rokhsar DS, Banfield JF (2004) Community structure and metabolism through reconstruction of microbial genomes from the environment. *Nature*, **428**(6978), 37–43.

Uehara H, Iwasaki Y, Wada C, Ikemura T, Abe T (2011) A novel bioinformatics strategy for searching industrially useful genome resources from metagenomic sequence libraries. *Genes & Genetic Systems*, **86**(1), 53–66.

Van Mooy BAS, Fredricks HF (2010) Bacterial and eukaryotic intact polar lipids in the eastern subtropical South Pacific: Water-column distribution, planktonic sources, and fatty acid composition. *Geochimica et Cosmochimica Acta*, **74**(22), 6499–6516.

Van Mooy BAS, Rocap G, Fredricks HF, Evans CT, Devol AH (2006) Sulfolipids dramatically decrease phosphorus demand by picocyanobacteria in oligotrophic marine environments. *Proceedings of the National Academy of Sciences of the United States of America*, **103**(23), 8607–8612.

Villanueva L, Bale N, Hopmans EC, Schouten S, Damsté JSS (2013) Diversity and distribution of a key sulpholipid biosynthetic gene in marine microbial assemblages. *Environmental Microbiology*,

Vogl Kajetan, Bryant DA (2011) Elucidation of the Biosynthetic Pathway for Okenone in *Thiodictyon* sp. CAD16 Leads to the Discovery of Two Novel Carotene Ketolases. *Journal of Biological Chemistry*, **286**(44), 38521–38532.

Vogl K, Bryant DA (2012) Biosynthesis of the biomarker okenone:  $\chi$ -ring formation. *Geobiology*, **10**(3), 205–215.

Weber M, Teeling H, Huang S, Waldmann J, Kassabgy M, Fuchs BM, Klindworth A, Klockow C, Wichels A, Gerds G, Amann R, Glöckner FO (2011) Practical application of self-organizing maps to interrelate biodiversity and functional data in NGS-based metagenomics. *The ISME Journal*, **5**(5), 918–928.

Weng FC, Su C-H, Hsu M-T, Wang T-Y, Tsai H-K, Wang D (2010) Reanalyze unassigned reads in Sanger based metagenomic data using conserved gene adjacency. *BMC Bioinformatics*, **11**(1), 565.

Woebken D, Teeling H, Wecker P, Dumitriu A, Kostadinov I, DeLong EF, Amann R, Glöckner FO (2007) Fosmids of novel marine Planctomycetes from the Namibian and Oregon coast upwelling systems and their cross-comparison with planctomycete genomes. *The ISME Journal*, **1**(5), 419–435.

Zhang L, Cui X, Schmitt K, Hubert R, Navidi W, Arnheim N (1992) Whole genome amplification from a single cell: implications for genetic analysis. *Proceedings of the National Academy of Sciences*, **89**(13), 5847–5851.

Zhu W, Lomsadze A, Borodovsky M (2010) Ab initio gene identification in metagenomic sequences. *Nucleic acids research*, **38**(12), e132.

# APPENDIX 1

## Compound Concentrations for Chapters 2 & 3

**Supplemental Table S1a.** Concentrations of all measured FAMES presented in Chapter 2, Figure 1a. Concentrations are in µg compound per mg total lipid extract; *n.d.*, not determined or below detection limits.

µg/mg TLE	7 m Fxn F	7 m Fxn G	8 m Fxn F	8 m Fxn G	Sed Fxn F	Sed Fxn G	Sho Fxn F	Sho Fxn G
C <sub>12:0</sub>	<i>n.d.</i>	<i>n.d.</i>	<i>n.d.</i>	<i>n.d.</i>	0.0712	0.0092	0.0046	<i>n.d.</i>
C <sub>13:0</sub>	<i>n.d.</i>	<i>n.d.</i>	<i>n.d.</i>	0.0025	<i>n.d.</i>	0.0044	0.0054	<i>n.d.</i>
C <sub>14:1</sub>	<i>n.d.</i>	0.0050	<i>n.d.</i>	0.0206	<i>n.d.</i>	0.0005	<i>n.d.</i>	<i>n.d.</i>
C <sub>14:0</sub>	1.49	0.120	0.859	0.753	0.286	0.0767	0.0728	0.0002
<i>i</i> -C <sub>15:0</sub>	0.0871	0.0238	0.269	0.634	0.136	0.0674	0.144	0.0020
<i>a</i> -C <sub>15:0</sub>	0.0471	0.0261	0.536	1.04	0.164	0.128	0.238	0.0056
C <sub>15:0</sub>	0.249	0.0188	0.0090	0.0883	0.0370	0.0137	0.0411	0.0001
<i>i</i> -C <sub>16:0</sub>	<i>n.d.</i>	<i>n.d.</i>	0.123	0.331	0.102	0.0746	0.123	0.0027
C <sub>16:1</sub>	10.2	0.908	1.12	0.845	0.0265	0.106	0.161	0.0014
C <sub>16:0</sub>	12.1	0.842	4.59	4.66	1.36	0.439	0.567	0.0223
10Me-C <sub>16:0</sub>	<i>n.d.</i>	<i>n.d.</i>	0.0530	0.322	<i>n.d.</i>	0.0389	0.0679	0.0023
<i>i</i> -C <sub>17:0</sub>	<i>n.d.</i>	<i>n.d.</i>	<i>n.d.</i>	0.0750	0.0385	0.0144	0.0620	0.0002
<i>a</i> -C <sub>17:0</sub>	<i>n.d.</i>	<i>n.d.</i>	0.130	0.376	0.0730	0.0525	0.133	0.0055
<i>cyc</i> -C <sub>17:0</sub>	<i>n.d.</i>	0.0894	<i>n.d.</i>	0.215	<i>n.d.</i>	<i>n.d.</i>	0.0137	<i>n.d.</i>
C <sub>17:0</sub>	0.396	0.0292	0.0499	0.123	0.0435	0.0234	0.0460	0.0002
<i>i</i> -C <sub>18:0</sub>	<i>n.d.</i>	<i>n.d.</i>	<i>n.d.</i>	0.0223	0.0041	0.0055	0.0075	<i>n.d.</i>
C <sub>18:2</sub>	0.129	0.0181	<i>n.d.</i>	0.357	0.0481	0.0042	0.0492	<i>n.d.</i>
C <sub>18:1</sub>	29.4	2.77	6.09	2.80	0.868	0.456	0.437	<i>n.d.</i>
C <sub>18:0</sub>	0.588	0.0541	1.60	1.53	0.882	0.203	0.264	0.0151
<i>i</i> -C <sub>19:0</sub>	<i>n.d.</i>	<i>n.d.</i>	<i>n.d.</i>	<i>n.d.</i>	<i>n.d.</i>	0.0086	0.0103	<i>n.d.</i>
<i>a</i> -C <sub>19:0</sub>	<i>n.d.</i>	<i>n.d.</i>	<i>n.d.</i>	<i>n.d.</i>	0.0036	0.0035	0.0060	<i>n.d.</i>
<i>cyc</i> -C <sub>19:0</sub>	<i>n.d.</i>	<i>n.d.</i>	<i>n.d.</i>	0.0951	0.0065	0.0326	0.0635	0.0004
C <sub>19:1</sub>	<i>n.d.</i>	<i>n.d.</i>	<i>n.d.</i>	<i>n.d.</i>	0.0136	<i>n.d.</i>	<i>n.d.</i>	<i>n.d.</i>
C <sub>19:0</sub>	<i>n.d.</i>	<i>n.d.</i>	<i>n.d.</i>	0.186	0.0032	0.0042	0.0173	0.0014
C <sub>20:1</sub>	0.645	0.0476	<i>n.d.</i>	0.0113	<i>n.d.</i>	0.0249	0.0287	<i>n.d.</i>
C <sub>20:0</sub>	<i>n.d.</i>	<i>n.d.</i>	0.101	0.111	0.269	0.0318	0.138	0.0032
C <sub>21:0</sub>	<i>n.d.</i>	<i>n.d.</i>	<i>n.d.</i>	<i>n.d.</i>	0.0251	<i>n.d.</i>	0.0278	<i>n.d.</i>
C <sub>22:0</sub>	<i>n.d.</i>	<i>n.d.</i>	0.0976	<i>n.d.</i>	1.24	0.0814	0.259	0.0015
C <sub>22:1</sub>	<i>n.d.</i>	<i>n.d.</i>	<i>n.d.</i>	<i>n.d.</i>	<i>n.d.</i>	<i>n.d.</i>	0.188	<i>n.d.</i>
C <sub>23:0</sub>	<i>n.d.</i>	<i>n.d.</i>	<i>n.d.</i>	<i>n.d.</i>	0.223	0.0207	0.0277	<i>n.d.</i>
C <sub>24:0</sub>	<i>n.d.</i>	<i>n.d.</i>	0.210	0.0520	3.23	0.0969	0.401	0.0026
C <sub>25:0</sub>	<i>n.d.</i>	<i>n.d.</i>	<i>n.d.</i>	<i>n.d.</i>	0.199	0.0077	0.0223	<i>n.d.</i>
C <sub>26:0</sub>	<i>n.d.</i>	<i>n.d.</i>	0.0714	0.0037	1.37	0.0253	0.132	<i>n.d.</i>
C <sub>27:0</sub>	<i>n.d.</i>	<i>n.d.</i>	<i>n.d.</i>	<i>n.d.</i>	0.0795	<i>n.d.</i>	<i>n.d.</i>	<i>n.d.</i>
C <sub>28:0</sub>	<i>n.d.</i>	<i>n.d.</i>	<i>n.d.</i>	<i>n.d.</i>	0.688	0.0078	0.0514	<i>n.d.</i>
C <sub>29:0</sub>	<i>n.d.</i>	<i>n.d.</i>	<i>n.d.</i>	<i>n.d.</i>	0.0440	<i>n.d.</i>	<i>n.d.</i>	<i>n.d.</i>
C <sub>30:0</sub>	<i>n.d.</i>	<i>n.d.</i>	<i>n.d.</i>	<i>n.d.</i>	0.316	0.0012	0.0348	<i>n.d.</i>
C <sub>32:0</sub>	<i>n.d.</i>	<i>n.d.</i>	<i>n.d.</i>	<i>n.d.</i>	0.0712	<i>n.d.</i>	0.0100	<i>n.d.</i>

**Supplemental Table S1b.** Carbon isotope values of all measured FAMES presented in Chapter 2, Figure 1b. Isotope values are in per mil; *n.d.*, not determined or below detection limits.

$\delta^{13}\text{C}$ (‰)	7 m Fxn F	7 m Fxn G	8 m Fxn F	8 m Fxn G	Sed Fxn F	Sed Fxn G	Sho Fxn F
<b>C<sub>14:0</sub></b>	-34.7 ± 0.3	<i>n.d.</i>	-30.3 ± 0.3	<i>n.d.</i>	-25.8 ± 0.3	-27.1 ± 0.5	-25.3 ± 0.2
<b><i>i</i>-C<sub>15:0</sub></b>	<i>n.d.</i>	<i>n.d.</i>	-32.4 ± 0.3	<i>n.d.</i>	-23.2 ± 0.3	-24.5 ± 0.5	-25.4 ± 0.2
<b><i>a</i>-C<sub>15:0</sub></b>	<i>n.d.</i>	<i>n.d.</i>	-32.8 ± 0.3	<i>n.d.</i>	-26.1 ± 0.3	-26.2 ± 0.2	-25.4 ± 0.2
<b>C<sub>15:0</sub></b>	<i>n.d.</i>	<i>n.d.</i>	<i>n.d.</i>	<i>n.d.</i>	<i>n.d.</i>	<i>n.d.</i>	-23.9 ± 0.5
<b><i>i</i>-C<sub>16:0</sub></b>	<i>n.d.</i>	<i>n.d.</i>	-31.5 ± 0.5	<i>n.d.</i>	-26.7 ± 0.5	-28.8 ± 0.5	-26.4 ± 0.2
<b>C<sub>16:1</sub></b>	-34.9 ± 0.3	-33.2 ± 0.3	-28.8 ± 0.3	-31.6 ± 0.5	<i>n.d.</i>	-24.9 ± 0.3	-23.8 ± 0.2
<b>C<sub>16:0</sub></b>	-36.4 ± 0.3	-34.0 ± 0.3	-33.8 ± 0.3	-37.4 ± 0.3	-28.3 ± 0.2	-28.2 ± 0.2	-28.4 ± 0.2
<b>10Me-C<sub>16:0</sub></b>	<i>n.d.</i>	<i>n.d.</i>	<i>n.d.</i>	<i>n.d.</i>	<i>n.d.</i>	-25.9 ± 0.5	-25.8 ± 0.2
<b><i>a</i>-C<sub>17:0</sub></b>	<i>n.d.</i>	<i>n.d.</i>	<i>n.d.</i>	<i>n.d.</i>	-27.9 ± 0.5	-27.1 ± 0.5	-26.6 ± 0.2
<b>C<sub>18:2+</sub></b>							
<b>Phytanic</b>	<i>n.d.</i>	<i>n.d.</i>	<i>n.d.</i>	<i>n.d.</i>	-25.6 ± 0.3	-23.3 ± 0.5	-30.5 ± 0.2
<b>C<sub>18:1</sub></b>	-31.4 ± 0.4	-31.8 ± 0.3	-33.5 ± 0.3	-31.6 ± 0.5	-27.2 ± 0.3	-24.7 ± 0.2	-25.2 ± 0.3
<b>C<sub>18:0</sub></b>	<i>n.d.</i>	<i>n.d.</i>	-27.4 ± 0.3	<i>n.d.</i>	-27.7 ± 0.2	-26.5 ± 0.2	-26.7 ± 0.2
<b>C<sub>20:0</sub></b>	<i>n.d.</i>	<i>n.d.</i>	<i>n.d.</i>	<i>n.d.</i>	-27.5 ± 0.3	<i>n.d.</i>	-30.8 ± 0.3
<b>C<sub>22:0</sub></b>	<i>n.d.</i>	<i>n.d.</i>	<i>n.d.</i>	<i>n.d.</i>	-27.2 ± 0.2	<i>n.d.</i>	-26.5 ± 0.3
<b>C<sub>22:1</sub></b>	<i>n.d.</i>	<i>n.d.</i>	<i>n.d.</i>	<i>n.d.</i>	<i>n.d.</i>	<i>n.d.</i>	-27.8 ± 0.2
<b>C<sub>23:0</sub></b>	<i>n.d.</i>	<i>n.d.</i>	<i>n.d.</i>	<i>n.d.</i>	-26.0 ± 0.3	<i>n.d.</i>	<i>n.d.</i>
<b>C<sub>24:0</sub></b>	<i>n.d.</i>	<i>n.d.</i>	<i>n.d.</i>	<i>n.d.</i>	-26.5 ± 0.2	<i>n.d.</i>	-23.0 ± 0.3
<b>C<sub>28:0</sub></b>	<i>n.d.</i>	<i>n.d.</i>	<i>n.d.</i>	<i>n.d.</i>	-25.0 ± 0.3	<i>n.d.</i>	<i>n.d.</i>

**Supplemental Table S2.** Concentrations of sterols presented in Chapter 2, Figure 3a. Concentrations are in  $\mu\text{g}$  compound per mg total lipid extract; *n.d.*, not determined or below detection limits.

$\mu\text{g}/\text{mg TLE}$	7m	8m	Sed	Sho
$\text{C}_{27}\Delta^{5,22}$	<i>n.d.</i>	2.556	6.018	<i>n.d.</i>
$\text{C}_{27}\Delta^{22}$	<i>n.d.</i>	2.527	2.532	0.179
$\text{C}_{27}\Delta^5$	0.117	19.64	43.36	0.301
$\text{C}_{27}$	0.002	10.91	10.47	0.931
$\text{C}_{28}\Delta^{5,22}$	<i>n.d.</i>	2.751	7.712	<i>n.d.</i>
$\text{C}_{28}\Delta^{22}$	<i>n.d.</i>	6.039	2.298	<i>n.d.</i>
$\text{C}_{28}\Delta^5$	<i>n.d.</i>	3.758	9.034	0.180
$\text{C}_{28}$	<i>n.d.</i>	1.815	0.930	<i>n.d.</i>
$\text{C}_{29}\Delta^{5,22}$	0.734	42.78	45.68	1.342
$\text{C}_{29}\Delta^{22}$	<i>n.d.</i>	<i>n.d.</i>	<i>n.d.</i>	0.419
$\text{C}_{29}\Delta^{5,25}$	<i>n.d.</i>	0.598	0.774	0.196
$\text{C}_{29}\Delta^5$	0.375	37.17	31.11	1.540
$\text{C}_{29}$	<i>n.d.</i>	6.927	4.255	1.118
$\text{C}_{29}\Delta^{5,24}$	<i>n.d.</i>	0.413	1.279	<i>n.d.</i>
$\text{C}_{29}\Delta^7$	<i>n.d.</i>	0.809	2.158	0.007

**Supplemental Table S3.** Concentrations of fatty alcohols presented in Chapter 2, Figure 2b. Concentrations are in  $\mu\text{g}$  compound per mg total lipid extract; *n.d.*, not determined or below detection limits.

$\mu\text{g}/\text{mg TLE}$	7 m	8 m	Sed	Sho
<i>n</i> -C <sub>16</sub>	0.020	0.137	0.026	0.008
<i>n</i> -C <sub>17</sub>	<i>n.d.</i>	0.005	<i>n.d.</i>	0.017
<i>n</i> -C <sub>18</sub>	0.078	0.443	0.112	0.028
<i>n</i> -C <sub>20</sub>	0.000	0.126	0.006	0.014
<i>n</i> -C <sub>22</sub>	<i>n.d.</i>	0.333	1.130	0.253
<i>n</i> -C <sub>24</sub>	0.004	0.338	1.840	0.509
<i>n</i> -C <sub>26</sub>	0.007	0.702	3.468	1.399
<i>n</i> -C <sub>28</sub>	<i>n.d.</i>	0.109	1.127	0.598
<i>n</i> -C <sub>30</sub>	<i>n.d.</i>	<i>n.d.</i>	0.204	0.183

**Supplemental Table S4.** Carotenoid Abundances in different Mahoney Lake samples. Values are in µg compound per mg total lipid extract.

	Acetone QQQ- MSMS	Acetone QQQ-MS	Acetone UV	Bligh-Dyer QQQ- MSMS	Bligh-Dyer Fxn MS	Bligh-Dyer Fxn UV
(a) 7-m						
<b>Bacteriopheophytin <i>a</i></b>	288.57	288.57	288.57	288.57	16.60	10.48
<b>Pyrobacteriopheophytin <i>a</i></b>	0.23	0.38		0.69	0.03	0.01
<b>Phaeophytin <i>a</i></b>	3.08	33.22	9.25			
<b>Thiothece-484</b>	0.03	3.82				
<b>Echinenone</b>		1.01				
<b>Okenone</b>	10.79	669.10	96.76	0.06	0.001	0.001
<b>Canthaxanthin</b>					0.02	
<b>Beta-carotene</b>			0.53			
<b>Lutein &amp; Zeaxanthin</b>				0.03		
	Acetone QQQ- MSMS	Acetone QQQ-MS	Acetone UV	Bligh-Dyer QQQ- MSMS	Bligh-Dyer Fxn MS	Bligh-Dyer Fxn UV
(b) Sediment						
<b>Bacteriopheophytin <i>a</i></b>	425.65	425.65	425.65	425.65	31.64	19.56
<b>Pyrobacteriopheophytin <i>a</i></b>	62.81	98.19	4.63	65.92	19.73	3.11
<b>Phaeophytin <i>a</i></b>	8.30	72.89	38.74		0.78	
<b>Thiothece-484</b>	0.90	72.23			3.93	0.20
<b>Echinenone</b>	0.11	50.00	14.92			
<b>Okenone</b>	61.17	3629.44	1212.88	139.57	14.13	1.43
<b>Canthaxanthin</b>	0.06	17.01	1.19		2.20	
<b>Beta-carotene</b>		0.00	59.22			
<b>Lutein &amp; Zeaxanthin</b>				2.63		
	Acetone QQQ- MSMS	Acetone QQQ-MS	Acetone UV	Bligh-Dyer QQQ- MSMS	Bligh-Dyer Fxn MS	Bligh-Dyer Fxn UV
(c) Shoreline						
<b>Bacteriopheophytin <i>a</i></b>	15.91	15.91	15.91	15.91	1.07	0.36
<b>Pyrobacteriopheophytin <i>a</i></b>	2.80	4.25	0.13	9.25	0.92	0.18
<b>Phaeophytin <i>a</i></b>	1.76	11.17	6.49			
<b>Thiothece-484</b>	0.27	17.91	5.05		0.34	0.02
<b>Echinenone</b>	0.01	5.29	1.56			
<b>Okenone</b>	0.22	13.77	5.68	0.67	0.35	0.03
<b>Canthaxanthin</b>	0.01	2.70	0.27		0.47	0.02
<b>Beta-carotene</b>			1.81			
<b>Lutein &amp; Zeaxanthin</b>				0.10		